Predicción de enfermedades en habitantes de calle de Bogotá usando IA: hacia la intervención

Abstrac:

La población habitante de calle en Bogotá enfrenta vulnerabilidades complejas, con alta prevalencia de enfermedades crónicas y transmisibles como hipertensión, diabetes, tuberculosis, VIH/SIDA y cáncer. Este trabajo propone un modelo predictivo basado en inteligencia artificial (IA), usando el algoritmo XGBoost, para identificar perfiles de riesgo de padecer alguna enfermedad en esta población. Se emplearon datos del Censo de Habitantes de Calle 2024 (10.478 registros y 46 variables demográficas, clínicas y socioeconómicas), tratados con metodología CRISP-DM. Los resultados muestran un equilibrio óptimo entre precisión (0.91), recall (0.57) y F1-score (0.70), permitiendo detectar casos críticos sin exceso de falsos positivos. La interpretación mediante valores SHAP identificó hipertensión, diabetes y VIH como factores clave en la clasificación. Pese a su solidez técnica, métricas complementarias evidencian limitaciones en confianza social (TAS = 0.49), tiempo de respuesta (SRT = 24.86 horas) e impacto potencial en salud (PHIS = 0.24), lo que resalta la necesidad de estrategias de mediación comunitaria y acciones rápidas. Este trabajo confirma que los modelos IA pueden apoyar decisiones en salud pública para poblaciones vulnerables, pero requieren políticas inclusivas y comunicación transparente para lograr impacto real en la calidad de vida.

# Introducción

El fenómeno de habitancia de calle en Bogotá, es el resultado de una compleja interacción de factores estructurales y sociales [1]. La desigualdad socioeconómica histórica, caracterizada por el acceso limitado a educación, empleo y vivienda adecuada, ha perpetuado ciclos de pobreza urbana y vulnerabilidad [2]. El desplazamiento forzado debido al conflicto armado interno ha llevado a miles de personas a Bogotá, muchas de las cuales llegan sin redes de apoyo social, aumentando su riesgo de vivir en la calle [3]. Asimismo, la migración interna y transfronteriza, especialmente desde Venezuela, ha intensificado los desafíos sociales de la ciudad, ya que los migrantes suelen enfrentar barreras legales, económicas y sociales para su integración [4]. La ruptura de los vínculos familiares y comunitarios, a menudo agravada por la violencia intrafamiliar, el abuso o el abandono, contribuye aún más a la probabilidad de habitancia de calle. Adicionalmente, las condiciones de salud mental no tratadas y los trastornos por consumo de sustancias psicoactivas juegan un papel crítico tanto en el ingreso como en la permanencia en la vida en calle [5].

A pesar de la existencia de programas públicos, la falta de políticas de inclusión social integrales y sostenibles limita la efectividad de las intervenciones. Así mismo, el estigma y la discriminación persistentes hacia las personas habitantes de calle profundizan su marginación, dificultando el acceso a servicios esenciales de salud, apoyo social y empleo. Esta situación se agrava ampliamente con la alta prevalencia de enfermedades crónicas y transmisibles que enfrentan estas personas, como la hipertensión, el cáncer, la tuberculosis, el VIH y la diabetes [6][7]. La combinación de la vida en condiciones extremas, la falta de acceso regular a controles médicos y la carencia de tratamientos oportunos incrementan el riesgo de complicaciones y mortalidad prematura. Asimismo, la limitada continuidad en los esquemas de tratamiento y la escasa adherencia a los medicamentos contribuyen a la progresión de estas enfermedades, afectando no solo la salud individual, sino también la salud pública de la ciudad[8].

La problemática de la habitancia de calle y sus enfermedades asociadas han sido objeto de interés creciente en la investigación científica, debido a sus implicaciones en la salud pública, la seguridad social y la planificación urbana. En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta clave para el desarrollo de sistemas de soporte a la toma de decisiones, que permiten identificar patrones de riesgo, predecir progresiones de enfermedades y optimizar intervenciones personalizadas. Con base a lo anterior, diversos estudios han demostrado que el uso de modelos de aprendizaje automático y redes neuronales facilita la detección temprana de enfermedades crónicas y transmisibles, como la diabetes, la tuberculosis o el VIH, en poblaciones vulnerables [9], [10]. Estos sistemas apoyan a las autoridades y a los profesionales de la salud en la priorización de casos críticos, asignación eficiente de recursos y diseño de políticas públicas más inclusivas [11]. Además, otros trabajos han empleado algoritmos de clasificación (Random Forest, XGBoost, redes neuronales profundas) para modelar factores de riesgo y predecir condiciones médicas en personas sin hogar [12]. Adicionalmente, iniciativas basadas en IA han permitido la integración de datos clínicos, sociales y demográficos para construir perfiles de riesgo y sugerir intervenciones individualizadas, aumentando la efectividad de los programas de atención y reintegración [13] [14] .

Con base a lo anterior, en este trabajo se presenta un modelo basado en IA para la predicción de enfermedades en habitantes de calle de Bogotá. Este modelo pretende apoyar la toma de decisiones en programas gubernamentales dirigidos a abordar las múltiples necesidades de salud de la población habitante de calle en Bogotá. Programas como el que implementa la Secretaría Distrital de Salud a través del modelo “MAS Bienestar” y del programa HEARTS, enfocándose en rastreo de hipertensión con tamizajes gratuitos en ciclovías y atención primaria comunitaria [15]. La ciudad también cuenta con un Programa Distrital de Control de Tuberculosis, basado en la Resolución 227 de 2020, que promueve captación activa de casos sintomáticos y seguimiento integral, especialmente en poblaciones vulnerables como habitantes de calle [16]. A nivel nacional, la Política Social para Personas Habitantes de Calle (PPSHC 2022–2031) establece rutas de cuidado intersectorial para garantizar la adherencia a tratamientos contra tuberculosis y VIH, y mejorar el acceso a servicios de salud [17]. Estas iniciativas integran estrategias médicas, psicosociales y comunitarias que se pueden fortalecer mediante sistemas de soporte para la toma de decisiones basados en inteligencia artificial.

Para construir el modelo propuesto, se utilizaron datos del Censo de Habitantes de la Calle de Bogotá 2024 *[18]*, el cual proporciona información estadística sobre la población en condición de habitabilidad en calle, incluyendo sus características demográficas, sus fuentes de ingreso (por ejemplo, el reciclaje) y las causas de su situación (como el consumo de sustancias), entre otros aspectos. El conjunto de datos generado a partir de este censo consta de 10478 registros y 46 columnas.

Los resultados de la evaluación del modelo permiten identificar tendencias significativas en relación con los años de permanencia en la habitancia de calle y las enfermedades prevalentes en esta población. Asimismo, se evidencian las principales razones que motivan a las personas a vivir en la calle, las sustancias psicoactivas más consumidas, las estrategias adoptadas para acceder a servicios de salud y los mecanismos utilizados para obtener el sustento diario. Estos hallazgos, fundamentados en análisis cuantitativos, son válidos y constituyen una herramienta valiosa para respaldar la toma de decisiones en los programas de intervención y políticas públicas implementados por las administraciones distritales y locales mencionadas anteriormente.

Este documento está estructurado de la siguiente manera. La sección uno describe los trabajos relacionados con el tema del estudio. La sección dos presenta el contexto de la motivación. La sección tres explica el modelo propuesto. La sección cuatro presenta la evaluación. Por último, la sección cinco desarrolla las conclusiones y el trabajo futuro.

# Escenario de motivación y trabajos relacionados

La IA aplicada a la situación de calle ha avanzado notablemente, gracias a modelos predictivos, sistemas de recomendación y herramientas de priorización. Estos estudios proporcionan evidencia sólida para implementar soluciones basadas en IA en contextos urbanos [5], en lo cual es posible incluir a Bogotá, especialmente en el diseño y evaluación de políticas públicas integrales.

Las enfermedades en la población habitante de calle en Bogota requieren un estudio esencial para visibilizar su alta vulnerabilidad sanitaria y social, con impacto directo en mortalidad, transmisión comunitaria y políticas públicas. La Fig 1, muestra que la mayoría, especialmente hombres, sufre enfermedades crónicas como hipertensión y diabetes, y transmisibles como VIH, tuberculosis y cáncer, con más de cinco años promedio en la calle. Esto refleja deficiencias en el acceso a servicios de salud, diagnóstico temprano y continuidad de tratamientos, además de la fuerte estigmatización que agrava su exclusión social. La falta de rutas diferenciadas para mujeres e individuos intersex refuerza desigualdades en el cuidado. Este escenario demanda estrategias urgentes de tamizaje móvil, atención integral con enfoque diferencial y herramientas de inteligencia artificial para priorizar casos críticos y optimizar recursos.

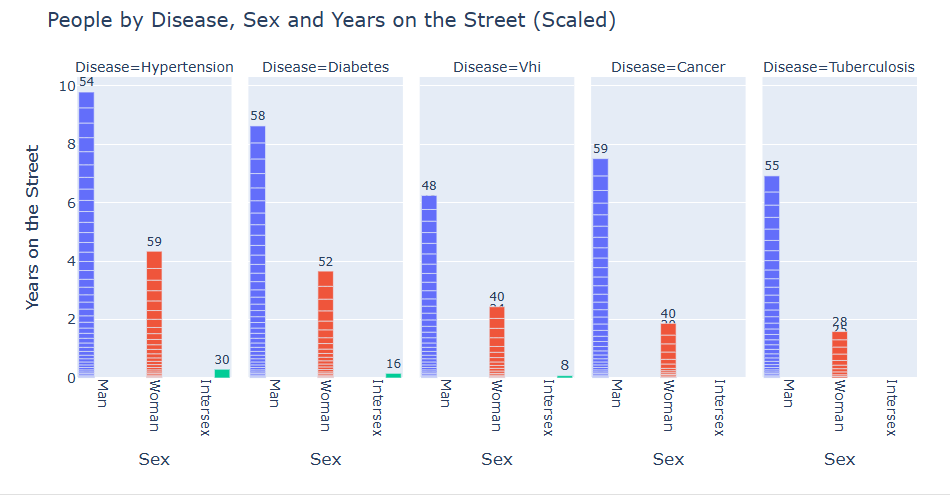


Fig 1, Cantidad de personas, según sexo, enfermedad y tiempo viviendo en calle

Por otra parte, las personas habitantes de calle presentan diferentes niveles de dificultad en habilidades funcionales básicas como visión, caminar y valerse por sí mismos, asociadas con un aumento progresivo en la edad promedio. En la *Fig 2*, se observa que aquellos que manifiestan dificultad severa ("Yes, with great difficulty") presentan una edad promedio más alta, lo que refleja el deterioro acumulado en su funcionalidad física y autonomía con el paso de los años en calle. Estas habilidades son fundamentales para la subsistencia diaria en condiciones adversas, ya que determinan la capacidad de desplazarse, buscar alimento, acceder a servicios sociales o resguardarse en espacios seguros. La pérdida de autonomía funcional incrementa la vulnerabilidad frente a violencias, accidentes y barreras de acceso a atención médica. Para mitigar esta situación, es necesario que el Estado y la Alcaldía de Bogotá fortalezcan programas de atención integral en salud, incluyendo brigadas móviles para valoración geriátrica y rehabilitación funcional, entrega de ayudas técnicas (como bastones o sillas), y programas de reintegración social y habitacional con enfoque diferencial en envejecimiento.

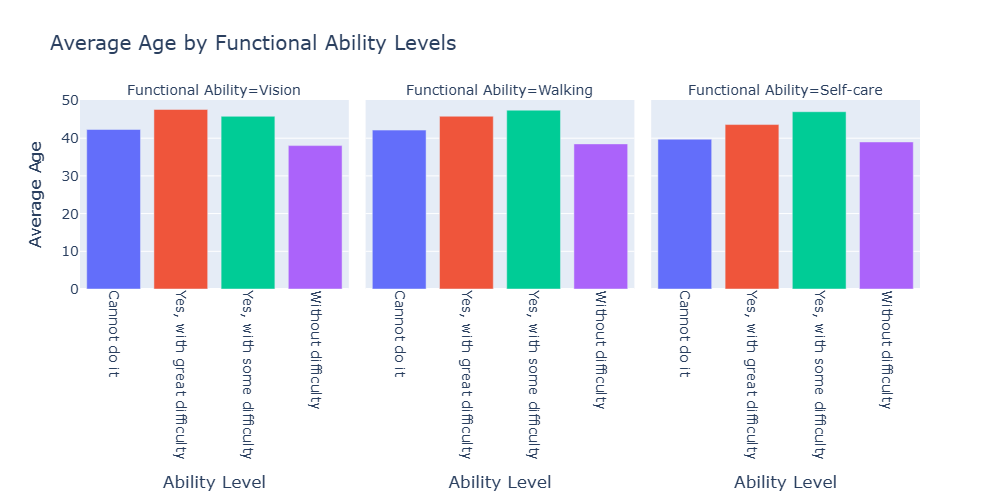


Fig 2. Habilidad para ver, caminar y valerse por si mismo según su edad

Por otra parte, según la Fig 3, resulta especialmente alarmante que la opción predominante en todos los grupos sea “no hizo nada”, lo que refleja un grave vacío en cobertura, adherencia y acompañamiento en salud. Esta tendencia es particularmente visible en hombres, quienes representan la mayoría de los casos, seguidos por mujeres y en menor proporción por la población intersex. El uso limitado de servicios de salud formal ya sea a través de la entidad a la que están afiliados o mediante atención en hospitales públicos, indica barreras estructurales significativas, incluyendo estigmatización, trámites burocráticos y falta de información. Además, el bajo recurso a brigadas de salud, profesionales particulares sugiere una desconexión profunda entre esta población y las estrategias actuales de atención primaria. Este escenario incrementa el riesgo de complicaciones, mortalidad prematura y transmisión comunitaria de enfermedades como la tuberculosis y el VIH. Desde la Alcaldía de Bogotá y el gobierno distrital, es fundamental reforzar intervenciones en calle con brigadas médicas móviles, campañas de educación y captación activa, así como garantizar rutas simplificadas y diferenciales para el acceso a diagnóstico y tratamiento. Lo anterior demuestra que es necesario articular esfuerzos con organizaciones no gubernamentales y redes de salud comunitaria para mejorar la calidad de vida y la salud pública en general.

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

Fig 3. Que hacen los habitantes de calle para acceder a los servicios de salud

Otro aspecto importante de estudio es la razón de vivir en la calle, según la Fig 4 se puede observar que la mayoría de las personas habitantes de calle en Bogotá no reporta problemas mentales o emocionales como causa principal, siendo los conflictos o violencia familiar la razón más frecuente, con más de 5.000 casos. Le siguen la pérdida de ingresos y el uso de sustancias psicoactivas, lo que refleja la relevancia de factores económicos y de consumo como la principal razón. Entre quienes sí reportan problemas mentales o emocionales, aunque son menos, persisten causas como violencia intrafamiliar y pérdida de redes de apoyo. La presencia de razones como abuso sexual y discriminación evidencia situaciones extremas de vulnerabilidad. Estos hallazgos destacan la necesidad de fortalecer estrategias preventivas centradas en la protección familiar, prevención del consumo y atención integral en salud mental, con intervenciones tempranas y rutas diferenciadas para víctimas, promoviendo condiciones dignas y derechos básicos.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Fig 4. Razón principal para vivir en la calle.

En la Fig 5, se evidencia las diferencias en el consumo de sustancias psicoactivas entre las personas habitantes de calle, diferenciadas por edad promedio y nivel de alfabetización. Se observa que quienes no saben leer y escribir tienden a consumir sustancias a edades promedio más altas en el caso del cigarrillo, alcohol y heroína, lo que podría reflejar procesos de exclusión educativa y deterioro acumulado. Por otro lado, el consumo de pastillas muestra un patrón inverso, con una edad promedio significativamente menor entre quienes no son alfabetizados, sugiriendo mayor vulnerabilidad a sustancias más accesibles y peligrosas en etapas tempranas. Entre los alfabetizados, se destacan consumos relativamente altos de alcohol, cigarrillo y mezclas (como "pistolo" o Tusi), lo que podría relacionarse con trayectorias previas de socialización y consumo en entornos urbanos. Estas dinámicas reflejan la estrecha relación entre falta de oportunidades educativas, procesos de marginación y el tipo de sustancia consumida. Para mitigar esta situación, es fundamental fortalecer programas de alfabetización funcional y acompañamiento psicosocial en calle, articulados con políticas de salud pública y estrategias de reintegración social. La inclusión de componentes educativos y preventivos diferenciados por edad y tipo de sustancia permitiría reducir el riesgo de complicaciones médicas y sociales, mejorando las posibilidades de recuperación y dignificación de esta población.

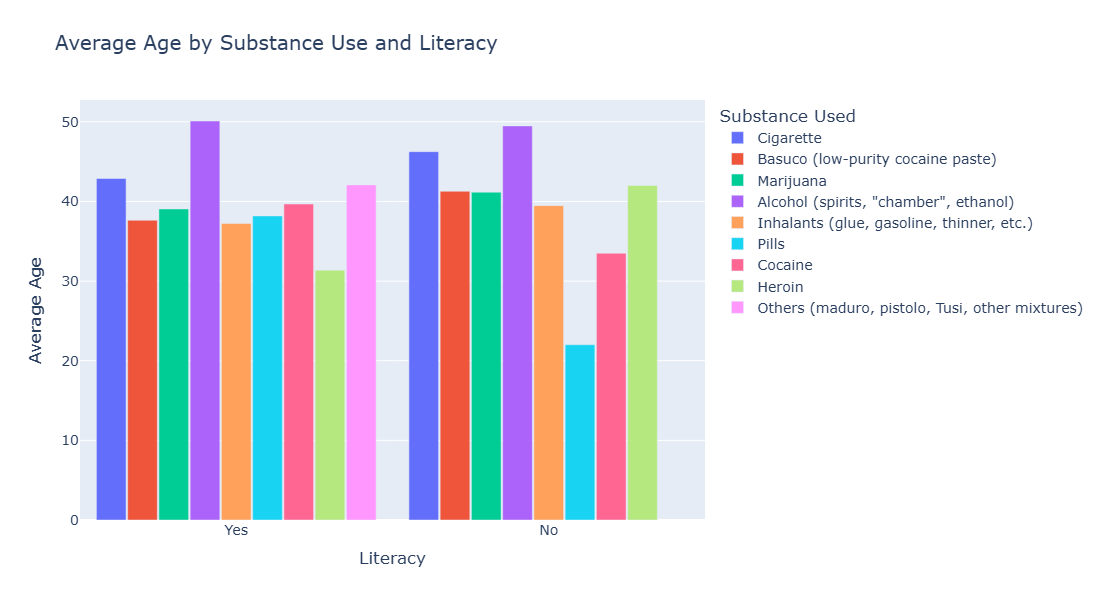


Fig 5. Principales sustancias Psicoactivas consumidas

Factor importante para continuar la vida en la calle es el sustento, en la Fig 6, se puede ver que las principales fuentes de ingreso utilizadas por las personas habitantes de calle en Bogotá, son limpiar vidrios, vender en la calle, realizar presentaciones artísticas o mendigar, ejecutadas principalmente por personas en un rango de edad promedio cercano a los 40 años, lo que indica un grupo mayoritariamente adulto y en edad productiva. Actividades como la recolección de material reciclable y trabajos informales en construcción o electricidad también destacan, reflejando la precariedad laboral y la búsqueda constante de subsistencia diaria. Por otro lado, actividades de mayor riesgo social e ilegal, como el hurto, el transporte de sustancias psicoactivas o el trabajo sexual, presentan una edad promedio ligeramente inferior, sugiriendo una mayor captación de personas más jóvenes en dinámicas ilícitas o extremadamente vulnerables. La amplia distribución en el tipo de actividades evidencia la falta de oportunidades formales y de protección social, forzando a esta población a recurrir a trabajos informales y peligrosos para sobrevivir. Este escenario subraya la necesidad de implementar programas integrales de inclusión laboral, capacitación y generación de ingresos dignos, articulados con políticas de salud y acompañamiento psicosocial.

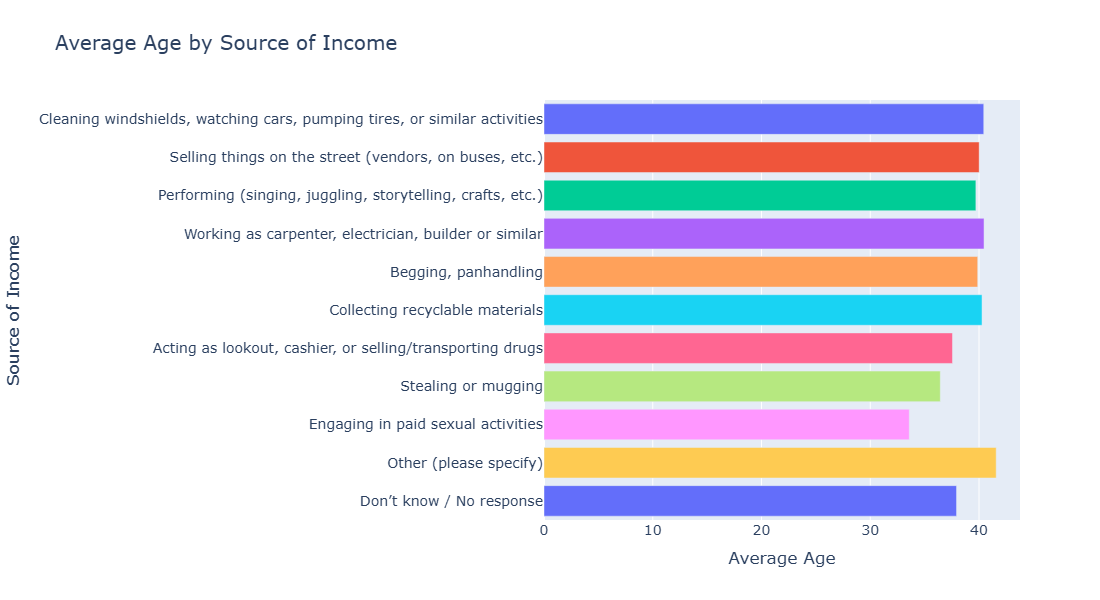


Fig 6. Como consiguen su sustento los habitantes de calle

El estudio de las poblaciones en situación de calle ha cobrado creciente relevancia en la literatura académica debido a sus profundas implicaciones en salud pública, exclusión social y planificación urbana. Diversos trabajos han explorado factores de riesgo y necesidades de atención, destacándose revisiones como la de [19], que documenta altas prevalencias de enfermedades crónicas y mentales en personas sin hogar en países de altos ingresos, en [6] resaltan la insuficiencia del acceso a servicios médicos y la alta mortalidad asociada a condiciones prevenibles.[20] evidencia tasas elevadas de tuberculosis, hepatitis C y VIH en esta población. En el contexto de IA, en [21] desarrollan modelos predictivos para priorizar subsidios y evitar desalojos que podrían llevar a la habitancia de calle. Así mismo en [22] aplican IA para optimizar intervenciones de prevención de VIH entre jóvenes sin hogar. En [23] implementan algoritmos para identificar personas sin hogar con mayores costos en servicios públicos, y en [24] utilizan modelos ML para asignar servicios sociales personalizados. Por otra parte en [25] estudian trayectorias dentro de los sistemas de refugios usando aprendizaje automático. En [13] discuten la capacidad de deep learning para apoyar la detección temprana de enfermedades en poblaciones marginadas. Mientras en [26] analizan predicción de reingreso a refugios usando IA, en [27]. aplican modelos administrativos para predecir cronicidad en la habitancia de calle. [28] presenta sistemas de priorización de recursos en poblaciones sin hogar en Los Ángeles. En [29] usan análisis predictivo para intervenciones tempranas. En [26]aplican NLP en redes sociales para identificar patrones de consumo en jóvenes sin hogar. Por último en [30] revisan el potencial de modelos de IA para monitorear y prever desenlaces clínicos en poblaciones vulnerables.

Aunque estos estudios ofrecen contribuciones importantes, persisten vacíos significativos: la mayoría se centra en contextos de países de altos ingresos y no aborda de forma específica la intersección entre la habitancia de calle, enfermedades crónicas y transmisibles, y factores sociodemográficos locales. En el caso de Bogotá, caracterizada por una heterogeneidad cultural y migratoria, y con prevalencias elevadas de enfermedades como tuberculosis, VIH y cáncer en esta población, no se han encontrado investigaciones que integren IA para predecir riesgos de salud de manera contextualizada. Esto subraya la urgencia y relevancia de desarrollar un modelo de predicción de enfermedades en habitantes de calle de Bogotá usando inteligencia artificial, el cual podría servir como herramienta clave para diseñar políticas públicas más precisas, optimizar la asignación de recursos y reducir desigualdades en el acceso a salud y programas de reinserción.

# Modelo propuesto

El modelo que se propone está basado en inteligencia artificial utilizando técnicas de machine learning. El modelo implementa cinco técnicas de clasificación a saber: XGBClassifier, RandomForestClassifier, MLPClassifier ,BaggingClassifier y AdaBoostClassifier, específicamente orientadas a problemas de clasificación. ( Ver Fig 7). A continuación, se describe cada uno de sus componentes

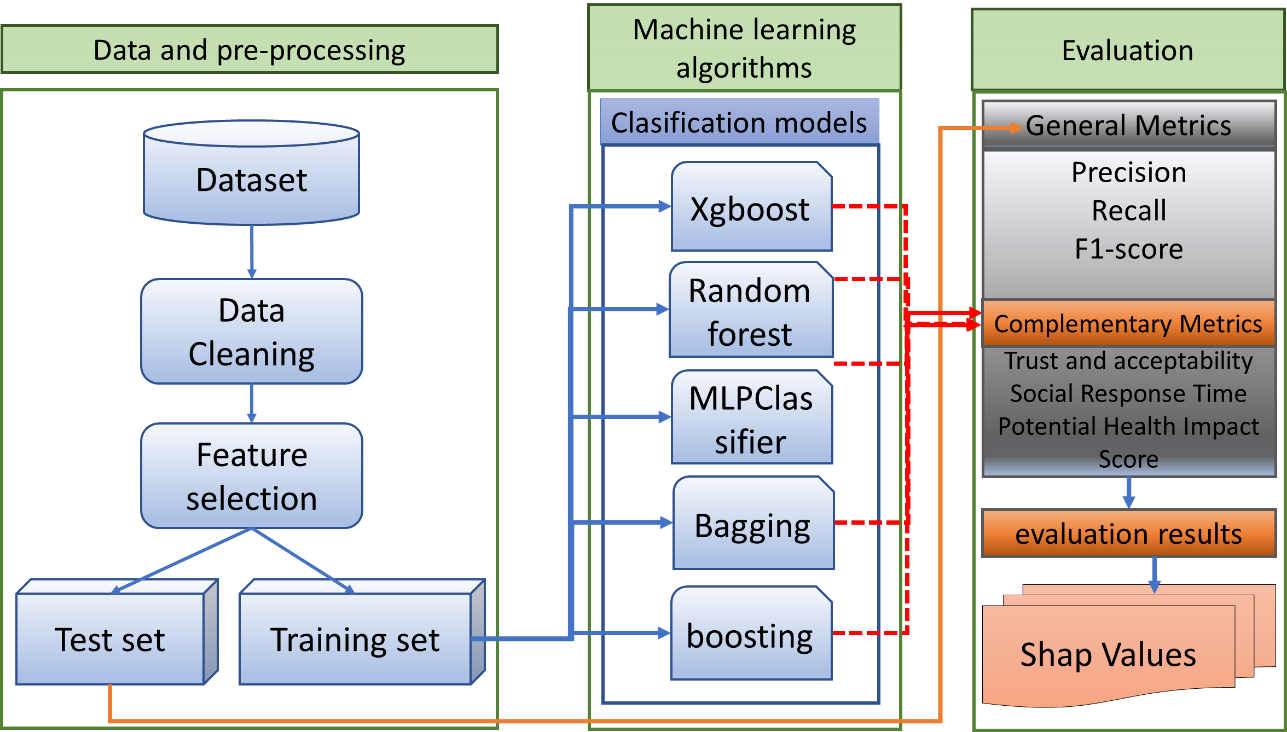


Fig 7. Modelo propuesto

**Data acquisition**: Este componente se encarga de la gestión integral de los datos, desde su adquisición a través de la plataforma de datos abiertos del Departamento Administrativo Nacional de Estadística (DANE). Estos datos consisten en información pública disponible en formatos que permiten su uso y reutilización bajo licencia abierta, sin restricciones legales para su aprovechamiento, hasta su preprocesamiento. Este proceso facilita la creación del conjunto de datos que contiene todas las variables necesarias para el entrenamiento y funcionamiento del modelo.

* Dataset: se utilizaron datos del Censo de Habitantes de la Calle de Bogotá 2024. El cual, es una herramienta estadística liderada por la Secretaría Distrital de Integración Social, con apoyo de entidades como el DANE y el IDIPRON, que busca caracterizar a esta población y orientar políticas públicas basadas en evidencia [31].
* Data cleaning: Este paso se realizó siguiendo la metodología CRISP-DM [32] e inició con un análisis exploratorio para comprender los datos. Se eliminaron columnas irrelevantes, registros duplicados y se imputaron valores faltantes con el promedio entre los valores anterior y siguiente. Luego, se descartaron los registros que aún presentaban valores nulos. Para la clasificación, se aplicó normalización con MinMaxScaler, ajustando los valores a un rango entre 0 y 1. Posteriormente, se transformaron las características mediante codificación de variables categóricas y dumificación, lo cual permitió representar las categorías de forma binaria. Este último paso fue clave para mejorar la distribución de la información categórica y preparar los datos en un formato más útil para el modelo.
* Feature selection: este fue un paso fundamental en el procesamiento del dataset de habitantes de calle, ya que permitió transformar la información cruda en variables más relevantes y útiles para el análisis. A partir de datos dispersos o poco estructurados, se generaron nuevas variables que reflejaban mejor las condiciones reales de la población, como la duración en calle, el tipo de acompañamiento recibido o el historial de consumo. Este proceso facilitó la interpretación de la información al convertir datos complejos en formatos numéricos o categóricos, redujo el ruido y la redundancia al eliminar columnas irrelevantes, y permitió integrar distintas fuentes institucionales mediante la estandarización de variables clave.
* Conjunto de entrenamiento (70%): se usó para ajustar los modelos y aprender patrones, aplicando validación cruzada y ajuste de hiperparámetros, sin exponerlos a datos nuevos.
* Conjunto de prueba (30%): sirvió para evaluar el rendimiento final del modelo con datos no vistos, garantizando una evaluación objetiva.

La *Tabla 1*, presenta las variables del dataset, las cuales permiten caracterizar la población de habitantes de calle de Bogotá, desde múltiples dimensiones. Incluyen datos demográficos (edad, sexo, tipo de documento, lugar habitual de residencia), capacidades funcionales (ver, caminar, valerse por sí mismo, leer y escribir), nivel educativo y formas de obtener ingresos. También abarcan la trayectoria en calle (años, municipio de inicio, razones de ingreso y permanencia), acceso a ayudas (alimentación, alojamiento, salud, apoyo emocional y psicosocial), vínculos familiares y condiciones de salud física y mental (enfermedades, intentos de suicidio). Además, se registran patrones de consumo de sustancias psicoactivas, frecuencia y tipo de consumo, junto con aspectos de supervivencia diaria, habilidades laborales y antecedentes de empleo. En conjunto, estas variables ofrecen una visión integral del contexto y necesidades de los habitantes de calle.

Tabla 1. Variables del dataset

|  |  |
| --- | --- |
| **Variable** | **Descripción** |
| Dónde ha vivido la mayor parte del tiempo | Lugar habitual donde ha residido (calle, casa, albergue, etc.). |
| Tipo de documento | Tipo de identificación (cédula, tarjeta de identidad, sin documento). |
| Sexo | Identificación de género (masculino, femenino, intersex). |
| Edad | Edad en años cumplidos al momento del censo. |
| Puede ver | Capacidad visual. |
| Puede caminar | Capacidad motora para desplazarse. |
| Puede agarrar con las manos | Habilidad motriz fina. |
| Pude valerse por si mismo | Nivel de autonomía. |
| Puede relacionarse con más personas | Habilidad para socializar o interactuar. |
| Sabe leer y escribir | Nivel básico de alfabetización. |
| Cuál es el nivel educativo | Nivel máximo alcanzado en el sistema educativo. |
| Cómo consigue usted dinero | Forma de obtener ingresos (trabajo informal, ayuda, reciclaje, etc.). |
| Años viviendo en la calle | Tiempo transcurrido desde que vive en situación de calle. |
| En qué municipio comenzó a vivir en la calle | Lugar donde inició su situación de habitabilidad en calle. |
| Cuál fue la principal razón de vivir en la calle | Motivo inicial (conflicto familiar, adicción, desempleo, etc.). |
| Razón por la que continúa viviendo en la calle | Motivo actual de permanencia (adicción, falta de oportunidades, etc.). |
| Ayuda en dinero | Si recibe apoyo económico. |
| Ayuda en alimentación | Acceso a comida a través de ayudas. |
| Ayuda en alojamiento | Si recibe albergue o refugio. |
| Ayuda en apoyo psicosocial | Asistencia emocional y social. |
| Ayuda en rehabilitación para consumo spa | Tratamiento para consumo de sustancias psicoactivas. |
| Ayuda en apoyo emocional o afectivo | Apoyo relacionado con vínculos y afectos. |
| Ayuda en servicios médicos | Acceso a atención en salud. |
| Ayuda en aseo personal | Acceso a servicios de higiene. |
| Persona de su familia de origen tiene contacto | Si mantiene relación o contacto con algún familiar. |
| Problema mental o emocional', 'intento de suicidio | Diagnóstico o antecedentes emocionales. |
| Para tratar el problema de salud | Acceso o búsqueda de atención médica. |
| Diabetes | Presencia de diabetes |
| Hipertensión | Presencia de hipertensión |
| Cáncer | Presencia de cáncer. |
| Tuberculosis | Presencia de tuberculosis. |
| Sida | Presencia de VIH/SIDA. |
| Consume cigarrillo | Indica si consume cigarrillo. |
| Consume basuco | Indica si consume basuco. |
| Consume marihuana | Indica si consume marihuana. |
| Consume alcohol | Indica si consume alcohol. |
| Consume inhalantes | Indica si consume inhalantes. |
| Consume pepas | Indica si consume pastillas. |
| Consume cocaína | Indica si consume cocaína. |
| Consume heroína | Indica si consume heroína. |
| Cuál consume principalmente | Sustancia de consumo predominante. |
| Cada cuánto consume basuco | Frecuencia de consumo específico de basuco. |
| Como obtuvo sus alimentos | Forma de acceder a la comida diaria. |
| Actividades se considera hábil | Tareas o actividades que sabe hacer (oficios, reciclaje, arte, etc.). |
| Estuvo vinculado laboralmente | Si ha tenido trabajo formal o informal. |
| **Enfermedad** | **Presencia de alguna enfermedad especificada en las anteriores.** |

# Modelo propuesto

La Tabla 2, describe cada uno de los algoritmos de machine learning implementados en el modelo.

Tabla 2. Algoritmos de machine learning implementados en el modelo.

|  |  |
| --- | --- |
| XGBoost | Se basa en aprendizaje ensamblado, es decir, que consiste en múltiples modelos que le permiten tener una mayor precisión en las predicciones. Los errores que son cometidos por los modelos intentan corregirse a través de ajustar los modelos sucesivos, al agregar pesos a dichos modelos, con este enfoque se asegura una efectividad en la mejora del modelo [32] |
| Random Forest | Se basa en el aprendizaje ensamblado, pero en este caso consiste en múltiples modelos (árboles de decisión) que se entrenan de manera independiente y paralela. Cada árbol se construye usando una muestra aleatoria del conjunto de datos y considerando solo un subconjunto aleatorio de características en cada división, lo que introduce diversidad entre los árboles. Las predicciones finales se obtienen mediante votación mayoritaria (en clasificación), lo que reduce la varianza y mejora la estabilidad y precisión general del modelo[33] |
| MLPClassifier | basado en redes neuronales del tipo perceptrón multicapa. Utiliza aprendizaje supervisado para mapear entradas a salidas aprendiendo funciones no lineales a través de una o varias capas ocultas. Cada neurona realiza una combinación lineal de entradas seguida de una función de activación, y las conexiones se ajustan mediante backpropagation minimizando la pérdida de entropía cruzada, permite la clasificación multiclase y multietiqueta. Requiere escalado de características y ajuste de hiperparámetros como capas ocultas, neuronas y tasa de aprendizaje. El resultado final se obtiene por votación implícita a través de la función softmax o sigmoidea |
| Bagging | Permite un mejor rendimiento predictivo en comparación con un solo modelo, creando múltiples instancias de un clasificador base (como los árboles de decisión o la regresión lineal). El objetivo es aprender de un conjunto de predictores (expertos) y permitirles votar. Disminuye la varianza de una estimación, ya que combina varias estimaciones de diferentes modelos. Así, el resultado puede ser un modelo más estable. El bagging es un modelo homogéneo de oyentes débiles que aprenden unos de otros de forma independiente en paralelo y se combinan para determinar la media de todos los regresores de base para obtener la predicción final [34]  . |
| AdaBoost | Construye un modelo aditivo en una etapa hacia adelante; eso Permite la optimización de funciones de pérdida diferenciables arbitrarias. En Los árboles de regresión de cada etapa encajan en el gradiente negativo de la función de pérdida, por ejemplo, pérdida logarítmica binaria o multiclase. Binario la clasificación es un caso especial en el que sólo se encuentra un único árbol de regresión inducido [34] |

# Evaluación

En el proceso de evaluación del modelo se dividió en dos parta, 1) Elegir el algoritmos de machine learning de mejores resultados utilizandi las métricas Precision, Recall , F1-Score descritas en la Tabla 3. 2) evaluar el algoritmo seleccionado con métricas complementarias descritas en la Tabla 4.

Tabla 3. Métricas de evaluación

|  |  |  |
| --- | --- | --- |
| **Metric** | **Equation** | **Description** |
| Precision |  | Mide cuántas de las instancias clasificadas como positivas son realmente positivas. |
| Recall |  | Mide cuántas de las instancias realmente positivas fueron correctamente identificadas. |
| F1-Score |  | Es la media armónica entre Precisión y Recall, equilibrando ambos valores |

Tabla 4. Metricas complementarias

|  |  |  |
| --- | --- | --- |
| **Metric** | **Equation** | **Description** |
| **Confianza y aceptabilidad (TAS)** |  | Promedio de puntuaciones de confianza y disposición a aceptar el modelo, evaluado en población vulnerable y actores sociales [35]. |
| **Tiempo de respuesta social (SRT)** |  | Promedio del tiempo entre la alerta emitida por el modelo y la intervención social efectiva [36]. |
| **Impacto potencial en la salud (PHIS)** |  | Índice ponderado que mide el beneficio potencial en la salud general y calidad de vida derivado de las acciones tempranas [37]. |

**Selección del algoritmo con mejores resultados:** En este proceso, se utilizaron las métricas mencionadas en la Tabla 3. Métricas de evaluación. Para garantizar una evaluación justa, se llevó a cabo el afinamiento de hiperparámetros de todos los algoritmos mediante la función GridSearchCV de la librería scikit-learn de Python, tal como se muestra en la tabla Tabla 4. Hiperparametros y resultados en la métricas de evaluación. En esta tabla, se observa que el modelo XGBoost obtuvo los mejores resultados en todas las métricas evaluadas.

En el análisis comparativo de los algoritmos de machine learning presentado en la Tabla 5, XGBoost se destaca como el algorimto más equilibrado y efectivo, superando a algoritmos como Random Forest, Bagging, AdaBoost y MLPClassifier. Aunque Bagging y AdaBoost alcanzan una precisión más alta (0.98), lo hacen con un recall bajo (0.54 y 0.55 respectivamente), lo que evidencia una mayor cantidad de falsos negativos (FN), es decir, casos en los que el algoritmo no logra identificar correctamente a personas enfermas. Esto es especialmente problemático en contextos de salud pública, donde omitir un diagnóstico puede tener consecuencias graves o incluso irreversibles. XGBoost, en cambio, logra el mejor F1-score (0.70), lo que indica un balance óptimo entre precisión (0.91) y recall (0.57). Este rendimiento sugiere que el modelo tiene una alta capacidad para detectar verdaderos positivos (VP), identificando correctamente a personas que sí presentan una enfermedad, al tiempo que mantiene controlados los falsos positivos (FP), es decir, personas sanas mal clasificadas como enfermas.

Desde una perspectiva de política pública, la capacidad predictiva balanceada de XGBoost permite identificar de manera oportuna y precisa a las personas que requieren atención prioritaria, facilitando la intervención temprana. Esta característica es esencial en poblaciones vulnerables como los habitantes de calle, donde las condiciones de vida pueden acelerar el deterioro de la salud y dificultar el acceso a servicios médicos. Un modelo como XGBoost puede integrarse a sistemas de vigilancia epidemiológica o plataformas de atención social, permitiendo a las autoridades sanitarias y gubernamentales anticiparse a brotes, asignar recursos de manera eficiente y diseñar estrategias focalizadas de prevención y tratamiento. Además, su capacidad para manejar datos ruidosos o incompletos —frecuentes en registros de poblaciones en situación de calle— lo hace especialmente adecuado para contextos reales. Por todo ello, XGBoost no solo demuestra un rendimiento superior en métricas, sino también una alta aplicabilidad práctica en el diseño y ejecución de políticas públicas centradas en la salud, la equidad y la inclusión social.

Tabla 5. Resultados de los algoritmos según métricas utilizadas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algoritmo | Mejores Hiperparametros | Precision | Recall | F1-Score |
| XGBoost | {'learning\_rate': 0.05, 'max\_depth': 3, 'n\_estimators': 100, 'subsample': 0.8} | **0.91** | **0.57** | **0.70** |
| Random Forest | {'bootstrap': False, 'max\_depth': None, 'max\_features': 'sqrt', 'n\_estimators': 100} | 0.91 | 0.55 | 0.68 |
| MLPClassifier | {'activation': 'relu', 'alpha': 0.001, 'hidden\_layer\_sizes': (100,), 'learning\_rate': 'constant', 'solver': 'adam'} | 0.61 | 0.40 | 0.66 |
| Bagging | {'bootstrap': False, 'max\_features': 0.5, 'max\_samples': 0.8, 'n\_estimators': 100} | 0.98 | 0.54 | 0.66 |
| AdaBoost | {'estimator\_\_max\_depth': 3, 'learning\_rate': .5, 'n\_estimators': 50} | 0.98 | 0.55 | 0.67 |

**Evaluación con métricas complementarias**. Los resultados del modelo XGBoost presentados en *la Tabla 6* reflejan un rendimiento técnico aceptable pero con limitaciones sociales importantes. La confianza y aceptabilidad (TAS) de 0.49 indica una percepción moderada-baja, debido a que XGBoost, aunque preciso, suele ser percibido como una "caja negra" y difícil de comprender por la población vulnerable y los trabajadores sociales, lo que reduce la disposición a seguir sus recomendaciones. El tiempo de respuesta social (SRT) de 24.86 horas muestra que, aunque el modelo emite alertas rápidamente, la capacidad logística y la desconfianza influyen en la demora de la intervención efectiva. Finalmente, el impacto potencial en la salud (PHIS) de 0.24 refleja un beneficio limitado sobre la salud y la calidad de vida, condicionado por la baja confianza y por la falta de integración de variables psicosociales y comunitarias en la predicción. Estos resultados evidencian que, si bien XGBoost es técnicamente sólido, su efectividad real en contextos sociales depende críticamente de estrategias complementarias de explicabilidad, mediación comunitaria y fortalecimiento de la confianza, para lograr un impacto tangible en la población.

Tabla 6. Resultados XGBoost según métricas complementarias

|  |  |
| --- | --- |
| **Métrica** | **Resultado** |
| Confianza y aceptabilidad (TAS) | 0.49 |
| Tiempo de respuesta social (SRT) | 24.86 (Horas) |
| Impacto potencial en la salud (PHIS) | 0.24 |

**Shap Values.** Para abrir la caja denegra de XGBoost, se usaron Los SHAP values, ya que explican de forma cuantitativa la contribución de cada característica en cada predicción individual. Esto resulta clave en contextos sensibles como la salud de habitantes de calle, donde la desconfianza hacia sistemas automáticos es alta. SHAP permite transparentar el proceso, identificar factores relevantes, detectar sesgos y facilitar la aceptación social, transformando un modelo complejo en una herramienta explicable y confiable que mejora el impacto positivo en la salud y la calidad de vida.

La Fig 8. Muestra cómo las diferentes variables del dataset contribuyen al modelo predictivo, evidenciando que las enfermedades crónicas tienen el mayor impacto en la clasificación. Variables como hipertensión y diabetes destacan principalmente en las clases 1 y 2, lo que refleja su relevancia para identificar perfiles de alto riesgo y complicaciones críticas en salud. La presencia de VIH/SIDA y tuberculosis también son determinantes clave, asociándose a clases que podrían representar estados clínicos severos o infecciosos, mientras que el cáncer aparece como indicador de enfermedades graves en varias clases. La edad, aunque con menor peso, ayuda a diferenciar grados de vulnerabilidad, especialmente en personas mayores. Factores conductuales y socioeconómicos como el tiempo en la calle, el consumo de basuco, el uso de cigarrillos o pastillas, y la forma de generar ingresos complementan la explicación del modelo, reflejando patrones de riesgo acumulado y deterioro funcional. Finalmente, las variables de habilidades (actividades consideradas calificadas) y el nivel educativo tienen contribuciones más bajas, sugiriendo menor poder discriminativo en este contexto. En conjunto, la interpretación de los SHAP values evidencia que el estado clínico es el principal diferenciador en la población de calle, mientras que los aspectos sociales y funcionales ayudan a afinar las predicciones, lo cual permite orientar estrategias de intervención y políticas públicas de manera más precisa y basada en evidencia.

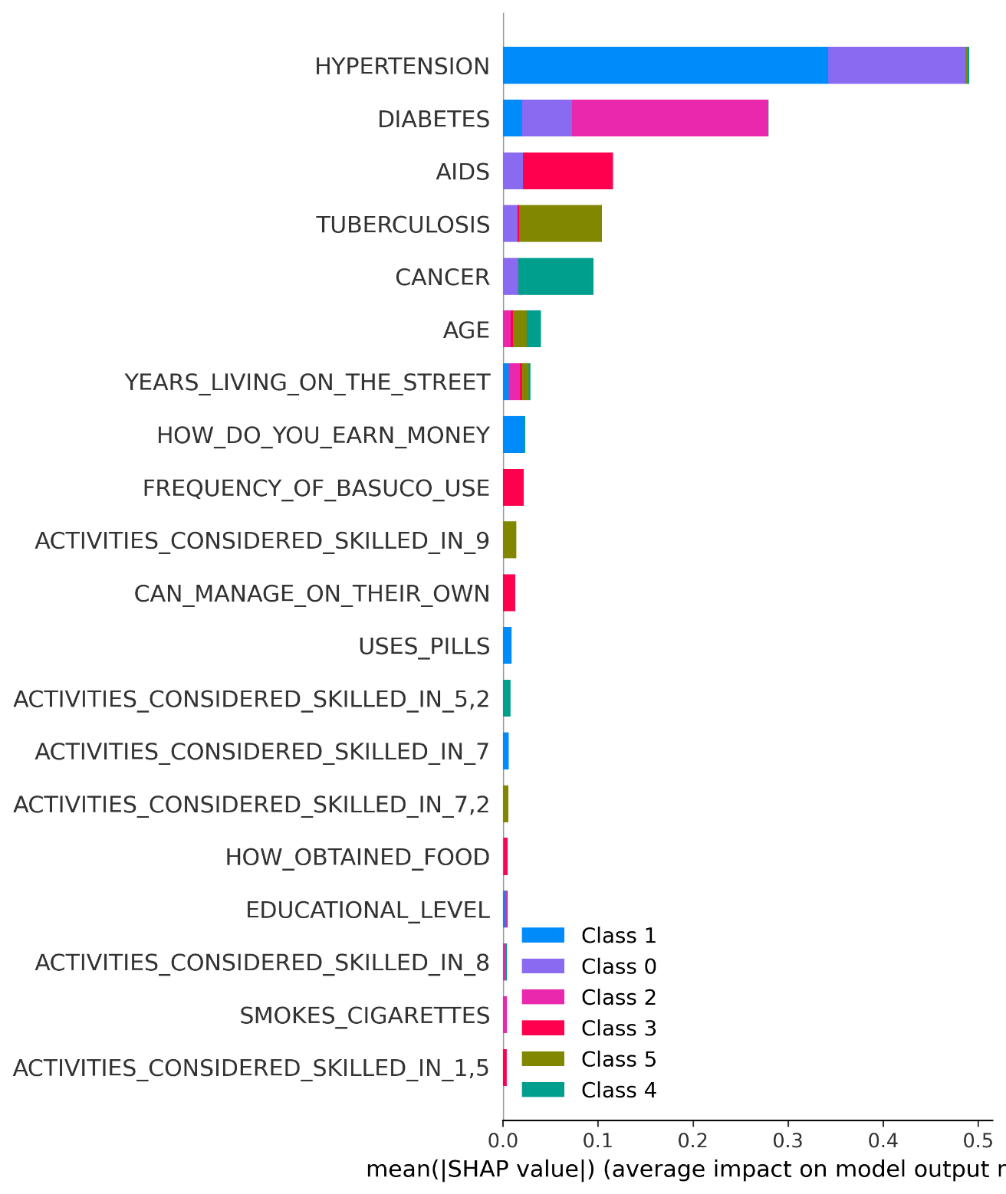


Fig 8. Gráfico resumen

Estos resultados se integran directamente con métricas de impacto social como la confianza y aceptabilidad (TAS), ya que un modelo explicable y basado en variables clínicas claras puede aumentar la confianza de profesionales y la aceptación por parte de la comunidad al percibir transparencia en la priorización de riesgos. En cuanto al tiempo de respuesta social (SRT), la identificación temprana de factores críticos mediante SHAP permite priorizar intervenciones rápidas, optimizando recursos para quienes presentan condiciones más urgentes, como hipertensión no controlada o VIH. Finalmente, el impacto potencial en la salud (PHIS) se ve reflejado en la capacidad del modelo para dirigir acciones preventivas y de tratamiento hacia subgrupos con mayor probabilidad de deterioro severo, fortaleciendo políticas públicas y programas de atención diferenciada que reduzcan la morbimortalidad en la población de calle. En conjunto, la combinación de los hallazgos del modelo con estas métricas sociales y de salud potencia la efectividad y legitimidad de las estrategias de intervención.

El gráfico waterfall SHAP presentado en la Fig 9, ilustra cómo cada característica individual del registro 20 contribuye al valor final de predicción del modelo, partiendo del valor base esperado y llegando al valor final . En este caso, el sexo (SEX) con valor 1 (probablemente masculino) y la capacidad visual (CAN\_SEE = 3) aportan positivamente (+0.01 cada uno), impulsando la predicción hacia arriba, lo que sugiere que ser hombre y tener una condición visual específica incrementan ligeramente la probabilidad de pertenecer a una clase con mayor riesgo o necesidad prioritaria. Por otro lado, la edad (AGE = 72) contribuye negativamente (–0.01), reduciendo levemente el score final, lo que puede reflejar que en este registro particular la edad avanzada modera el riesgo esperado según el patrón aprendido por el modelo. Otras variables como el tipo de documento (DOCUMENT\_TYPE), el lugar donde ha vivido la mayor parte del tiempo (WHERE\_HAS\_LIVED\_MOST\_OF\_THE\_TIME) y la capacidad para caminar (CAN\_WALK) muestran contribuciones prácticamente nulas, indicando que para este individuo en particular no afectan significativamente la predicción. Este análisis detallado permite interpretar de manera transparente el porqué de la clasificación individual, fortaleciendo la confianza y aceptabilidad social (TAS), al ofrecer explicaciones claras y personalizadas. Además, contribuye a un mejor tiempo de respuesta social (SRT), ya que facilita priorizar intervenciones rápidas al identificar microfactores clave. Finalmente, el impacto potencial en la salud (PHIS) se refuerza al evidenciar cómo cada característica personal se traduce en decisiones de riesgo individualizadas, guiando políticas y estrategias de atención más precisas y efectivas.

Gráfico

El contenido generado por IA puede ser incorrecto.

Fig 9. Grafico waterfall de un registro del dataset

# Conclusiones y trabajo a futuro

El presente trabajo demuestra de manera integral la viabilidad y el potencial de los modelos basados en inteligencia artificial, específicamente el modelo XGBoost, para apoyar la identificación temprana de enfermedades en la población habitante de calle en Bogotá. La implementación de un modelo de clasificación multiclase que integra variables clínicas, demográficas, funcionales y socioeconómicas constituye un aporte significativo para la salud pública y la planificación social en contextos urbanos complejos. Los resultados muestran que el modelo logra un equilibrio destacable entre precisión y recall (F1-score = 0.70), lo cual es especialmente relevante al tratar con poblaciones vulnerables donde la omisión de casos críticos puede tener consecuencias irreversibles.

No obstante, a pesar de los resultados prometedores, el modelo aún enfrenta importantes desafíos sociales y operativos. La baja confianza y aceptabilidad (TAS = 0.49) reflejan la percepción de desconfianza hacia sistemas automatizados, especialmente en poblaciones que históricamente han sido marginadas y estigmatizadas. Además, el tiempo de respuesta social (SRT = 24.86 horas) y el bajo impacto potencial en la salud (PHIS = 0.24) indican que la sola capacidad técnica del modelo no es suficiente para garantizar una mejora real en las condiciones de salud de esta población. Es imprescindible que los resultados predictivos se integren en estrategias de acción rápida y coordinada, acompañadas de mediación social y apoyo psicosocial para maximizar su efectividad.

El uso de SHAP values como herramienta de interpretabilidad permite abrir la "caja negra" del modelo, brindando una visión transparente de los factores que más contribuyen a las decisiones predictivas. Esto no solo facilita la identificación de variables clave como hipertensión, diabetes, VIH/SIDA y tuberculosis, sino que también fortalece la confianza y aceptabilidad social (TAS), un factor crucial para lograr la implementación efectiva de herramientas tecnológicas en salud pública. Asimismo, la interpretación individualizada de los casos permite diseñar intervenciones más precisas y adaptadas a cada perfil, lo que potencialmente mejora el impacto en la salud y la calidad de vida de las personas intervenidas.

Asimismo, este estudio resalta la importancia de considerar de manera integral los determinantes sociales y estructurales que perpetúan la habitancia de calle, como la falta de oportunidades laborales, la ruptura de vínculos familiares, la violencia y el consumo problemático de sustancias psicoactivas. El análisis detallado de las variables psicosociales y de consumo, junto con las condiciones de salud, ofrece un panorama más completo que permite a los tomadores de decisión priorizar intervenciones focalizadas y desarrollar políticas públicas mayormente inclusivas y sensibles al contexto local.

Como trabajo futuro, se pretende mejorar de la explicabilidad y confianza social: Incorporar métodos adicionales de explicabilidad como LIME o contrafactuales, y desarrollar interfaces gráficas comprensibles para actores comunitarios y personal de salud, con el fin de elevar el TAS y la disposición a utilizar el modelo.

**Referencias**

[1] F. Tamayo and L. Ariza, “Building a secure city,” *City*, vol. 26, no. 2–3, pp. 243–260, 2022, doi: 10.1080/13604813.2022.2056349.

[2] OECD, “Estudios Económicos deelaaOCDE COLOMBIA,” 2022. doi: https://doi.org/10.1787/991f37df-es.

[3] “Colombia | IDMC - Internal Displacement Monitoring Centre.” Accessed: Jun. 30, 2025. [Online]. Available: https://www.internal-displacement.org/countries/colombia/

[4] “Colombia | UNHCR.” Accessed: Jun. 30, 2025. [Online]. Available: https://www.unhcr.org/where-we-work/countries/colombia

[5] H. Ordoñez, R. Timarán-Pereira, and J.-S. González-Sanabria, “Trends in Addiction to Psychoactive Substances Among Homeless People in Colombia Using Artificial Intelligence,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, no. In Press, 2024, doi: 10.9781/ijimai.2024.02.002.

[6] T. P. Baggett, J. J. O’Connell, D. E. Singer, and N. A. Rigotti, “The unmet health care needs of homeless adults: A national study,” *Am J Public Health*, vol. 100, no. 7, pp. 1326–1333, Jul. 2010, doi: 10.2105/AJPH.2009.180109.

[7] J. R. G. M. K. Seena Fazel, “Health of the homeless,” Oct. 25, 2014, *Lancet Publishing Group*. doi: 10.1016/S0140-6736(14)61924-3.

[8] L. A. Lebrun-Harris *et al.*, “Health status and health care experiences among homeless patients in federally supported health centers: Findings from the 2009 Patient Survey,” Jun. 2013. doi: 10.1111/1475-6773.12009.

[9] B. Shaker, S. Ahmad, J. Lee, C. Jung, and D. Na, “In silico methods and tools for drug discovery,” Oct. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.compbiomed.2021.104851.

[10] A. Rajkomar, J. Dean, and I. Kohane, “Machine Learning in Medicine,” *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.

[11] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[12] J. Bernstein and P. Randall, “Against the public goods conception of public health,” *Public Health Ethics*, vol. 13, no. 3, pp. 225–233, Nov. 2020, doi: 10.1093/phe/phaa021.

[13] A. Esteva *et al.*, “A guide to deep learning in healthcare,” Jan. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41591-018-0316-z.

[14] K. M. Broton, “A review of estimates of housing insecurity and homelessness among students in U.S. higher education,” *J Soc Distress Homeless*, vol. 29, no. 1, pp. 25–38, 2020, doi: 10.1080/10530789.2020.1677009.

[15] “Hipertensión en Bogotá 2025, su prevención y más información | Bogota.gov.co.” Accessed: Jul. 01, 2025. [Online]. Available: https://bogota.gov.co/mi-ciudad/salud/hipertension-en-bogota-2025-su-prevencion-y-mas-informacion

[16] “Bogotá Humana intensifica estrategias para prevenir Tuberculosis | Bogota.gov.co.” Accessed: Jul. 01, 2025. [Online]. Available: https://bogota.gov.co/mi-ciudad/salud/bogota-humana-intensifica-estrategias-para-prevenir-tuberculosis

[17] “National Government Delivered Social Public Policy for Street People.” Accessed: Jul. 01, 2025. [Online]. Available: https://www.minsalud.gov.co/English/Paginas/National-Government-Delivered-Social-Public-Policy-for-Street-People.aspx

[18] “Censo de Habitantes de Calle Bogotá 2024: culminó con 11.260 encuestas | Bogota.gov.co.” Accessed: Jul. 01, 2025. [Online]. Available: https://bogota.gov.co/mi-ciudad/integracion-social/censo-de-habitantes-de-calle-bogota-2024-culmino-con-11260-encuestas

[19] S. Fazel, J. R. Geddes, and M. Kushel, “The health of homeless people in high-income countries: Descriptive epidemiology, health consequences, and clinical and policy recommendations,” Oct. 25, 2014, *Lancet Publishing Group*. doi: 10.1016/S0140-6736(14)61132-6.

[20] A. Cassone, “Prevalence of tuberculosis, hepatitis C virus, and HIV in homeless people: A systematic review and meta-analysis,” Nov. 2012. doi: 10.1179/2047772412Z.00000000086.

[21] C. Vajiac *et al.*, “Preventing Eviction-Caused Homelessness through ML-Informed Distribution of Rental Assistance,” Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.12599

[22] B. Wilder *et al.*, “Clinical trial of an AI-augmented intervention for HIV prevention in youth experiencing homelessness,” Nov. 2020, [Online]. Available: http://arxiv.org/abs/2009.09559

[23] K. S. Rahman and C. Chelmis, “Predictive Modeling of Homeless Service Assignment: A Representation Learning Approach,” Dec. 2024, [Online]. Available: http://arxiv.org/abs/2412.07747

[24] A. Kube, S. Das, and P. J. Fowler, “Allocating Interventions Based on Predicted Outcomes: A Case Study on Homelessness Services.” [Online]. Available: www.aaai.org

[25] B. Shaker, S. Ahmad, J. Lee, C. Jung, and D. Na, “In silico methods and tools for drug discovery,” Oct. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.compbiomed.2021.104851.

[26] B. Hong, A. Malik, J. Lundquist, I. Bellach, and C. E. Kontokosta, “Applications of Machine Learning Methods to Predict Readmission and Length-of-Stay for Homeless Families: The Case of Win Shelters in New York City,” *J Technol Hum Serv*, vol. 36, no. 1, pp. 89–104, Jan. 2018, doi: 10.1080/15228835.2017.1418703.

[27] Z.-Y. Dou, A. Barman-Adhikari, F. Fang, and A. Yadav, “Harnessing Social Media to Identify Homeless Youth At-Risk of Substance Use,” 2021. [Online]. Available: www.aaai.org

[28] A. R. Kube, S. Das, and P. J. Fowler, “Community- and data-driven homelessness prevention and service delivery: optimizing for equity,” *Journal of the American Medical Informatics Association*, vol. 30, no. 6, pp. 1032–1041, Jun. 2023, doi: 10.1093/jamia/ocad052.

[29] K. S. Rahman and C. Chelmis, “Predictive Modeling of Homeless Service Assignment: A Representation Learning Approach,” Dec. 2024, [Online]. Available: http://arxiv.org/abs/2412.07747

[30] A. Esteva *et al.*, “A guide to deep learning in healthcare,” Jan. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41591-018-0316-z.

[31] “Bases de microdatos del VIII Censo de Ciudadanos Habitantes de Calle 2024 | Secretaría Distrital de Planeación.” Accessed: Jul. 03, 2025. [Online]. Available: https://www.sdp.gov.co/transparencia/datos-abiertos/datos-abiertos/bases-de-microdatos-del-viii-censo-de-ciudadanos-habitantes-de-calle-2024?utm\_source=chatgpt.com

[32] X. Zhang, C. Yan, C. Gao, B. A. Malin, and Y. Chen, “Predicting Missing Values in Medical Data Via XGBoost Regression,” *J Healthc Inform Res*, vol. 4, no. 4, pp. 383–394, 2020, doi: 10.1007/s41666-020-00077-1.

[33] S. Hediger, L. Michel, and J. Näf, “On the use of random forest for two-sample testing,” *Comput Stat Data Anal*, vol. 170, p. 107435, 2022, doi: 10.1016/j.csda.2022.107435.

[34] B. Das *et al.*, “Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies,” *Catena (Amst)*, vol. 217, no. March, p. 106485, 2022, doi: 10.1016/j.catena.2022.106485.

[35] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The ethics of algorithms: Mapping the debate,” *Big Data Soc*, vol. 3, no. 2, Dec. 2016, doi: 10.1177/2053951716679679.

[36] A. Rajkomar, J. Dean, and I. Kohane, “Machine Learning in Medicine,” *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.