

Modelo predictivo para la identificación de factores de desigualdad social asociados a la discapacidad física en personas mayores

Información sobre los autores:

Luisa Fernanda Panesso Cardona, socióloga, estudiante de la Maestría en Gestión Estratégica en Información, Docente del Departamento de Antropología y Sociología de la Universidad de Caldas, Manizales, Colombia, luisa.panesso@ucaldas.edu.co; <https://orcid.org/0000-0002-5839-1580>.

Declaro que en la presente investigación no han incidido intereses o valores distintos a los que usualmente presentan las investigaciones.

Diego Samir Melo Solarte, Ingeniero de Sistemas, Especialista en telecomunicaciones, Magister en Ciencias de la Computación, Doctor en Desarrollo Sostenible; Docente de la Facultad de Ciencias e Ingeniería de la Universidad de Manizales; mdiego@umanizales.edu.co; <https://orcid.org/0000-0003-0941-6697>.

Declaro que en la presente investigación no han incidido intereses o valores distintos a los que usualmente presentan las investigaciones.

Resumen

El objetivo de esta investigación es analizar la pertinencia de un modelo predictivo para la identificación de factores de desigualdad social asociados a la discapacidad física en personas mayores en Colombia. Predecir la incidencia de factores contextuales en la prevalencia de la discapacidad en adultos mayores, permite potenciar el desarrollo de políticas públicas coherentes con las necesidades básicas de las poblaciones y con el diseño de estrategias de promoción y prevención de la discapacidad física. Se utilizó datos transversales de 23.694 adultos mayores en la encuesta nacional SABE Colombia de zonas rurales y urbanas para la construcción de un modelo predictivo de clasificación de árbol de decisión. Se analizaron los determinantes estructurales del envejecimiento saludable relacionadas con características sociodemográficas y capacidad intrínseca. La discapacidad se asoció con determinantes estructurales sociodemográficos como la edad avanzada y bajo nivel educativo, los resultados demuestran la necesidad de predecir en poblaciones envejecidas el riesgo de prevalencia de discapacidad en personas mayores a partir de características sociodemográficas y el índice de Barthel.

Palabras clave: Ciencia de datos; minería de datos; envejecimiento saludable; personas discapacitadas; desigualdad.

Abstract

The objective of this research is to analyze the relevance of a predictive model for the identification of social inequality factors associated with physical disability in older adults in Colombia. Predicting the incidence of contextual factors in the prevalence of disability in older adults allows enhancing the development of public policies consistent with the basic needs of the populations and with the design of strategies for the promotion and prevention of physical disability. Cross-sectional data from 23,694 older adults in the national SABE Colombia survey from rural and urban areas were used to construct a predictive decision tree classification model. Structural determinants of healthy aging related to sociodemographic characteristics and intrinsic capacity were analyzed. Disability was associated with sociodemographic structural determinants such as advanced age and low educational level. The results demonstrate the need to predict the risk of disability prevalence in older people in aging populations based on sociodemographic characteristics and the Barthel index.

Keywords: data science; data mining; healthy aging; disabled persons; inequality.

Introducción

En este estudio fue importante partir de la situación contextual de Colombia al ser uno de los países más desiguales de América Latina, hecho que incide en la calidad de vida de las personas mayores debido al acceso inequitativo a recursos valorados como consecuencia de la concentración de la riqueza, para el año 2021 el coeficiente de Gini fue de 0,523; en cuanto a la pobreza monetaria el 39,3% del total de la población se ubicó por debajo de la línea de pobreza, fenómeno que se agudiza en zonas rurales con el 44,6% de los hogares en condición de pobreza, mientras en zonas urbanas el 37,8% de los hogares son pobres. En lo concerniente a la medición de la pobreza de forma directa, expresada a través del índice de pobreza multidimensional, el 16% de la población se encontraba para el año 2021 en situación de pobreza multidimensional (DANE, 2021), el analfabetismo, el bajo logro educativo, el desempleo de larga duración y el empleo informal son las privaciones con mayor incidencia, las anteriores características adquiridas se relacionan con la prevalencia de la discapacidad física en personas mayores como consecuencia de los bajos niveles de formación académica, bajos niveles de ingresos y la necesidad de trabajar en la informalidad.

El acelerado proceso de envejecimiento que vive la población mundial, en especial los países en desarrollo, ha centrado la atención de gobiernos y organizaciones a nivel internacional desde hace varias décadas para generar estrategias, como la Década del Envejecimiento Saludable 2020-2030, que permitan afrontar las transformaciones en la composición de la estructura poblacional en un contexto de desigualdad social, en donde determinantes estructurales sociodemográficos como la edad avanzada, el sexo femenino, la residencia rural, el estado civil, vivir solo, bajo nivel de ingresos, el bajo nivel educativo y la pertenencia étnica repercuten en la prevalencia de discapacidad en personas mayores (Gómez, Osorio García, Panesso, & Curcio, 2021).

En el actual contexto de transformación social, la analítica de datos y la combinación de áreas de conocimiento, técnicas y herramientas como la minería

de datos y el aprendizaje automático han contribuido a las investigaciones en el área de la salud en la predicción del riesgo de padecer discapacidad de movilidad a partir de la velocidad de la marcha (Womens Hlth Initiative Investigato, 2018), que han conducido a propuestas como la inclusión de medidas de velocidad de la marcha en la atención geriátrica de rutina (Rosso, y otros, 2019). En relación a la capacidad funcional de las personas mayores, se evidenció la relación entre la mala condición de salud, el aumento de caídas, la disminución del cumplimiento de los servicios preventivos en salud y aumento de la utilización y los gastos en atención médica (Musich, Wang, Ruiz, Hawkins, & Wicker, 2018). La construcción del índice de fragilidad como herramienta clínica y de investigación arrojó como resultado que la falta información en dominios clínicos afecta la predicción del riesgo de discapacidad (Shi, McCarthy, Mitchell, & Kim, 2020). Por la línea del envejecimiento saludable y las afectaciones que genera la sarcopenia, se resaltó la importancia del análisis de las actividades de la vida diaria (ADV) y las actividades instrumentales (AIVD) al ser esenciales para la vida independiente de las personas mayores, en donde las medidas musculares son predictoras de futuras discapacidades y dependencia en la población mayor de 60 años (Wang, Yao, Zirek, Reijnierse, & Maier, 2020).

En el contexto internacional las investigaciones que se realizan alrededor del envejecimiento saludable se enfocan en la predicción de discapacidades físicas, pero no logran la convergencia entre el factor clínico de la discapacidad y la situación contextual de países como Colombia que se caracterizan por la desigualdad social. La pertinencia de la construcción de un modelo predictivo de discapacidad que considere variables contextuales, permitiría potenciar mediante el modelo de salud de promoción y prevención de la salud, iniciativas que favorezcan la satisfacción de necesidades básicas, la movilidad, el establecimiento y mantenimiento de relaciones sociales, la toma de decisiones, la contribución de los adultos mayores a la sociedad, el potenciar capacidades físicas y mentales de

los individuos y el diseño, implementación y ejecución de políticas públicas que generen entornos protectores durante el curso de vida.

Debido a la existencia de un conjunto de condiciones sociales de desigualdad que devienen en la estructuración de condiciones diferenciales, cada una de ellas aumentando las afectaciones en términos de discapacidad física que repercuten de forma negativa en el proceso de envejecimiento saludable de la población, esto conduce al siguiente cuestionamiento: ¿Cuál es el modelo predictivo, mediante la ciencia de datos, que permita la identificación de los factores de desigualdad social asociados a la discapacidad física en adultos mayores en Colombia, en relación con los determinantes del envejecimiento saludable?

Referentes conceptuales

La Cuarta Revolución Industrial se caracteriza por la adopción de las Tecnologías de Información y Comunicación (TICs), término acuñado en el año 2011 en el desarrollo de la Feria Tecnológica de Hannover-Messe, describe el proceso de digitalización de sistemas y procesos industriales (Joyanes Aguilar, 2017). Esta denominación evidencia el proceso de generación, almacenamiento y transformación de los datos que se ha venido presentando durante las últimas décadas, suceso que favorece la disponibilidad de grandes volúmenes de información y que invita a los investigadores a encontrar nuevas maneras para gestionarlos y aprovecharlos.

Ciencia de datos

Pensar en el aprovechamiento de los datos disponibles implica considerar el *Proceso de Extracción del Conocimiento a partir de Datos* (KDD, del inglés *Knowledge Discovery in Databases*). Gregory Piatetsky-Shapiro, acuñó este primer concepto debido al interés de ofertar un taller sobre el descubrimiento de datos, el cual buscaba superar el uso peyorativo del concepto minería de datos que se venía presentando a finales de la década de los años 80's. Logró cumplir su

objetivo y a partir de este primer taller realizado en la IJCAI-89, del inglés *International Joint Conference on Artificial Intelligence*, la industria de las tecnologías alcanzó considerables progresos en cuanto al avance técnico para la construcción del hardware más potente que apoyara el desarrollo de la *minería de datos*, no obstante, la comprensión holística de todo el proceso de descubrimiento de conocimiento a partir de los datos, abarca un conjunto de momentos que pasa por la adquisición de los datos, la limpieza y depuración de los datos útiles, por el entendimiento y procesamiento, pasando posteriormente a una fase de descubrimiento e interpretación de resultados que finalmente deben ser explotados en utilidad a las necesidades y posibilidades de una organización (Piatetsky-Shapiro , 2000).

La ciencia de datos es un concepto muy popular en la actualidad, su génesis es interdisciplinar, expertos coinciden en la convergencia de elementos de las matemáticas aplicadas, la informática, la consultoría empresarial y el desarrollo de nuevos productos (How to Develop Data Science Skills, 2015), si bien recoge un conjunto de principios fundamentales que sustentan y guían la extracción de información (Provost & Fawcett, 2013) y que conducen a la posibilidad de preprocesar los datos de tal forma que pueda realizarse procesamientos matemáticos y hacer un trabajo de análisis estadístico (Cady, 2017), para ello puede involucrarse metodologías, como CRISP-DM del inglés *Cross-Industry Standard Process for Data Mining*, al ofrecer técnicas y herramientas para adquirir, organizar, analizar y obtener conocimiento de grandes conjuntos de datos con el objetivo de tomar decisiones informadas y contribuir a la resolución de problemáticas de diversa índole (Hastie, Tibshirani, & Friedman, 2008).

El concepto de ciencia de datos, a su vez, se encuentra estrechamente relacionado con la minería de datos que consiste en resolver problemas analizando datos ya presentes en las bases de datos y debido al actual contexto en el que se da un asombroso aumento de las posibilidades de encontrar patrones en los datos al apoyarse en el conocimiento estadístico para su procesamiento y

de esta forma descubrir áreas de interés, patrones estructurales y posibles relaciones entre variables de estudio y dar lugar a nuevos conocimientos, o en el ámbito comercial, generar ventajas competitivas (Witten, Frank, & Hall, 2011). Los patrones útiles permiten hacer predicciones sobre nuevos conjuntos de datos, lo cual conduce hacia el aprendizaje automático, el cual puede considerarse un subcampo de la inteligencia artificial y permite que un programa informático extraiga conclusiones de un conjunto de datos basándose en la inferencia de datos (Earnshaw, Dill, & Kasik, 2019). Es así como la minería de datos proporciona técnicas para encontrar y describir patrones estructurales en los datos y el aprendizaje automático contribuye en la explicación y predicción del comportamiento de nuevos conjuntos de datos.

La naturaleza interdisciplinaria de la ciencia de datos aunada a las características de la minería de datos y del aprendizaje automático permitiría impulsar la toma de decisiones objetivas en problemáticas sociales que requiere una mirada multidimensional que direccionen el accionar de la salud pública alrededor del envejecimiento al englobar aportes de áreas del conocimiento como la sociología, la gerontología, la geriatría, la epidemiología, la psicología, entre otras, para consolidar estrategias preventivas, como es el caso de la presente propuesta, alrededor de los procesos de envejecimiento poblacional y la discapacidad física de personas mayores.

Envejecimiento saludable y desigualdad social

Desde el plano biológico, el envejecimiento se asocia con la acumulación de una gran variedad de daños moleculares y celulares, que con el pasar de los años reducen de forma gradual las reservas fisiológicas, con lo cual se incrementa el riesgo de padecimiento de enfermedades crónicas y se presenta una disminución en la capacidad general del individuo (Organización Mundial de la Salud, 2015). Sin embargo, el envejecimiento no sólo se asocia a estas pérdidas biológicas, también está relacionado con cambios sociales en cuanto se genera

una transformación en las relaciones sociales de las personas mayores y su rol dentro de un determinado entorno, es importante mencionar que el proceso de envejecimiento varía de un individuo a otro, no en todos los individuos se manifiesta de igual forma y esto está relacionado con el curso de vida de los mismos, por ello, desde la Organización Mundial de la Salud (OMS) se propuso el concepto de envejecimiento saludable al ser, algo más que la ausencia de enfermedad, el proceso de desarrollar y mantener la capacidad funcional que permite el bienestar en la vejez (Organización Mundial de la Salud, 2015).

El envejecimiento saludable aborda tres dimensiones: en primer lugar está la capacidad funcional que permite a la satisfacción de necesidades básicas, la movilidad, el establecimiento y mantenimiento de relaciones sociales, la toma de decisiones y la contribución a la sociedad; en segundo lugar, la capacidad intrínseca que versa sobre la combinación de las capacidades físicas y mentales de los individuos, que pueden verse afectados por la prevalencia de discapacidades que se manifiestan en la falta de independencia en las actividades de la vida diaria, como caminar, pensar, ver oír y recordar; finalmente, el entorno construido, el hogar, las relaciones sociales, actitudes, valores, política de salud (Organización Mundial de la Salud, 2015).

El envejecimiento saludable al propender por el desarrollo humano de las personas mayores, se relaciona con la situación contextual de desigualdad, desde el enfoque de capacidades se logra la convergencia de la perspectiva económica y social de esta problemática. La desigualdad no se limita a la distribución inequitativa de recursos valorados como consecuencia de la concentración de la riqueza, sino que incluye la falta de oportunidades y privación de capacidades básicas a las que se enfrentan las personas para llevar una vida digna en el ejercicio de sus libertades, para alcanzar sus metas (Sen, 1999), situación que se agudiza con la prevalencia de discapacidades físicas en la vejez, al afectarse su capacidad intrínseca.

La propuesta teórica del sociólogo Göran Therborn va en sintonía con el enfoque de capacidades anteriormente descrito, al considerar la desigualdad como una negación de la posibilidad de desarrollo de las capacidades humanas que puede ser susceptible de adoptar diferentes formas y conlleva múltiples consecuencias debido al orden sociocultural que afecta la capacidad para funcionar como seres humanos (Therborn, 2015), postura que a su vez podría relacionarse con el enfoque de desigualdad acumulada, al conducir a la mortalidad prematura y prevalencia de enfermedades crónicas como consecuencia de sistemas sociales que contribuyen y generan desigualdad manifiesta durante el curso de vida como consecuencia de procesos demográficos y de desarrollo; desde esta perspectiva la desventaja a la que están expuestos los individuos durante su curso de vida es una exposición al riesgo, mientras que la ventaja incrementa la exposición a la oportunidad, de esta manera las trayectorias de los individuos están determinadas por la acumulación de riesgos, recursos disponibles y la acción humana (Ferraro & Shippee, 2009).

En síntesis, los referentes conceptuales planteados en la presente investigación evidencian la pertinencia de la relación dada por la ciencia de datos y los principios fundamentales presentados en el marco de la *Década del Envejecimiento Saludable*; gracias a su carácter interdisciplinar, esta relación propicia la convergencia de diversas áreas del conocimiento que abordan desde el plano biológico, clínico, sociológico, económico o demográfico las afectaciones que sufren las personas mayores en su capacidad intrínseca como consecuencia del contexto de desigualdad social en el que se encuentran.

Es así como la contribución de esta investigación se enfoca en el análisis de la pertinencia de la construcción de un modelo predictivo que permita identificar el riesgo de desigualdad en una persona mayor y que a su vez, contribuya en la toma de decisiones al momento de construir políticas públicas y estrategias de promoción y prevención de la salud para garantizar entornos protectores para las personas mayores con discapacidad física o en riesgo de padecerla.

Materiales y métodos

Esta es una investigación de enfoque cuantitativo se basa en la metodología (CRISP-DM), que permite a los proyectos que incorporan minería de datos, la comprensión del fenómeno de estudio y la construcción de modelos acordes a los objetivos trazados y al conocimiento de los datos disponibles, a través de seis fases generales: comprensión del área problemática, conocimiento de los datos, preparación de los datos, modelamiento, evaluación del modelo, y finalmente el despliegue o implementación. (Schröera, Kruse, & Jorge, 2021).

Los datos utilizados provienen de la encuesta Salud, Bienestar y Envejecimiento (SABE), conocida por ser la primera encuesta transversal longitudinal realizada durante los años 1999-2000 por la Organización Panamericana de la Salud (OPS) falta referencia, en las principales ciudades de siete países de América Latina y el Caribe; en el caso de Colombia, la aplicación del instrumento se adelantó entre los años 2014 y 2015 a nivel nacional, con cobertura geográfica por regiones¹ referencia si es posible.

Este estudio recopila información para proyectar las necesidades de atención en salud de las personas mayores, mejorar su participación y seguridad social, y planificar estrategias, ante el acelerado proceso de transición demográfica que atraviesa el país, que contribuyan en la calidad de vida mediante la implementación de políticas públicas y programas que mitiguen los efectos no deseados del envejecimiento en la región latinoamericana y del Caribe (Dirección de Epidemiología y Demografía, 2018).

La encuesta cuenta con el respaldo de las políticas públicas de orden nacional alrededor del envejecimiento y vejez tanto en sus versiones anteriores, 2007-2019, 2015-2024, como en su más reciente actualización, 2022-2031, a través de la cual se incluye una línea de acción orientada a la promoción de las

¹ Atlántica: Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Sucre, Magdalena y San Andrés; Oriental: Boyacá, Meta, Putumayo, Cundinamarca, Casanare, Amazonas, Guainía, Guaviare, Vaupés, Vichada, Meta; Central: Antioquia, Caquetá, Caldas, Quindío, Risaralda, Tolima, Huila; Pacífica: Valle, Cauca, Nariño y Chocó; Bogotá: Bogotá.

investigaciones relacionadas con la vejez y el envejecimiento, además establece la próxima aplicación de la encuesta SABE durante el año 2025 (Colombia, 2022).

Para el desarrollo de esta investigación, uno de los criterios de selección de la muestra de personas mayores se orientó por el método clínico de calificación del estado cognitivo, conocido como Mini Examen del Estado Mental de Folstein (MMSE), el punto de corte se estableció en 13 de 35 puntos, dado que un resultado inferior es el reflejo de un deterioro cognitivo grave (Folstein, Folstein , & McHugh, 1975). El objetivo de la encuesta era conocer, mediante una evaluación interdisciplinaria, la situación de las personas mayores en el marco de los determinantes del envejecimiento activo, el anterior marco teórico y conceptual adoptado por la OMS y desde el modelo de los determinantes sociales en salud (Ministerio de Salud y Protección Social - Departamento Administrativo de Ciencia Tecnología e Innovación, COLCIENCIAS, Universidad del Valle y Universidad de Caldas, 2016).

Fase 1. Comprensión del área problemática

La comprensión del área problemática está dada por el entendimiento de la transición demográfica que atraviesa la humanidad, la incidencia de las características sociodemográficas de las personas mayores en la calidad de vida; a partir del análisis de investigaciones relacionadas con la predicción de diferentes aspectos de la salud y la discapacidad física en personas mayores, estos estudios en general utilizaron modelos y herramientas predictivas para abordar temas como la discapacidad, el riesgo de caídas, la estancia hospitalaria, la fragilidad y la incidencia del estrés en la percepción de salud.

Los resultados de estos estudios muestran la utilidad de las herramientas predictivas en la identificación de factores riesgo asociados a la discapacidad física y la contribución en la toma de decisiones en el ámbito de la salud, no obstante, no se evidenció el abordaje de la discapacidad en relación a los factores contextuales que agudizan esta afectación en la capacidad intrínseca de las personas mayores. En coherencia con lo anteriormente expuesto, se consolidaron los referentes conceptuales para proporcionar un marco de interpretación a los

datos, esto con el fin de favorecer el aprovechamiento de los datos disponibles mediante la implementación de la ciencia de datos y el envejecimiento saludable.

Fase 2. Conocimiento de los datos

Para el desarrollo de esta investigación se depuró la base de datos de SABE, la cual cuenta con más de 405 preguntas organizadas en 13 secciones (ver **¡Error! No se encuentra el origen de la referencia.**). Debido al enfoque social del abordaje de la discapacidad física del objetivo de esta investigación, se procedió a identificar en las categorías contenían información relacionada con la identificación de los individuos, el contexto socioeconómico y la capacidad intrínseca en relación a los objetivos de investigación establecidos alrededor del diseño de un modelo predictivo basado en minería de datos, para la identificación de los factores de desigualdad social asociados a la discapacidad física en personas mayores en Colombia, en relación con los determinantes del envejecimiento saludable.

Cuadro 1. Diseño de instrumentos

Sección	Temas
Identificación	• Identificación de la vivienda • Identificación del encuestado – Parte I • Filtro (MMSE) a persona mayor • Filtro (MMSE) proxy a acompañante • Identificación del encuestado – Parte II • Etnia
Aspectos socioeconómicos	• Educación • Información familiar, redes de apoyo y soporte social • Historia laboral y ocupación • Ingresos • Gastos
Medioambiente físico	• Vivienda y entorno residencial • Uso, tendencia y disponibilidad de tecnología • Prácticas frente a salud ambiental • Servicios públicos • Transporte
Medioambiente social	• Aspectos culturales • Actividades sociales y participación • Movilidad territorial (migración y desplazamiento) • Espacio de vida (relación con el entorno e independencia)
Conducta	• Valoración nutricional (Mini-nutricional) • Consumo de alcohol • Consumo de cigarrillo • Sexualidad
Cognición y afecto	• Valoración cognoscitiva Minimental Test de Folstein • Quejas subjetivas de pérdida de memoria • Escala de Depresión Geriátrica – Yesavage
Funcionalidad	• Índice de Barthel (Actividades Básicas de la Vida Diaria) • Actividades instrumentales de la vida diaria • Limitaciones funcionales • Actividad Física – AVD Social
Condiciones médicas de salud	Circunstancias tempranas y condiciones adversas en la infancia • Autopercepción del estado de salud • Enfermedades no transmisibles • Signos y

	<ul style="list-style-type: none"> síntomas generales Medicamentos Pruebas de tamizaje y actividades preventivas Órganos de los sentidos (audición, visión y salud bucal) Caídas y temor a caer Maltrato, abuso y discriminación Uso y acceso a servicios Hospitalizaciones
Uso y acceso a servicios de salud Antropometría y valoración funcional	<ul style="list-style-type: none"> Amputaciones Medidas antropométricas (Peso, talla, circunferencia de cintura, circunferencia de pantorrilla, altura de rodilla y circunferencia braquial) Escala corta de valoración funcional SPPB (equilibrio estático y velocidad de la marcha)
Enlace submuestras estudio	<ul style="list-style-type: none"> Enlace a cuidador familiar Enlace a biomarcadores Enlace a submuestra SABE
Registros submuestra estudio SABE persona mayor	<ul style="list-style-type: none"> Toma de presión arterial y pulso Escala corta de valoración funcional SPPB (incorporarse de una silla) Fuerza de agarre
Datos de control	

Fuente: Elaboración propia a partir de Dirección de Epidemiología y Demografía, 2018

Es así como las variables seleccionadas corresponden a datos acerca de características biológicas de los individuos, como sexo, edad, prevalencia de morbilidad y discapacidad, y características adquiridas relacionadas con el nivel educativo, pertenencia étnica, y el componente socioeconómico. En este primer acercamiento con los datos se procedió con la carga de los mismos y con el cálculo de medidas de tendencia central, de frecuencias y medidas de distribución para de esta forma identificar los principales factores asociados con discapacidad física en personas mayores según los determinantes del envejecimiento saludable y analizar las condiciones de desigualdad a partir de la caracterización socioeconómica.

Fase 3. Preparación de los datos

Los datos se prepararon con el fin de adecuarlos a las técnicas de minería de datos que se emplearon sobre ellos, con lo cual se seleccionó el subconjunto con el que podría construirse el modelo, se limpiaron para mejorar su calidad, se transformaron los datos a partir de los existentes para lograr darles el formato requerido por la herramienta de modelado, como es el caso del árbol de decisión, la regresión logística, la red neuronal artificial y el bosque aleatorio:

1. El punto de partida en la construcción de los modelos fue la carga de las distintas librerías.
2. Se crearon copias del DataFrame para no afectar de forma radical la base de datos.
3. La variable clave en la construcción de los modelos se identificó como BARTHEL al permitir clasificar a los sujetos con la característica principal respecto a la prevalencia de la discapacidad física, por consiguiente, todos los modelos se construyeron en relación a esta variable.
4. Se eliminaron variables, como departamento, para simplificar el modelado, esto estuvo determinado por el análisis descriptivo de los datos, allí se hizo relevante incluir variables como la edad, los convivientes, los años de educación, los ingresos, la multimorbilidad, entre otros.
5. Se utilizó en el modelo de regresión logística la función `get_dummies` para convertir las columnas categóricas del DataFrame (excepto BARTHEL), en variables ficticias, dado que se requería datos numéricos en lugar de categóricos.
6. Se calcularon correlaciones, en caso de ser necesario, entre la variable de estudio y todas las demás variables, posteriormente se filtraron las correlaciones que eran mayores o iguales a 0.1 en valor absoluto o menores o iguales a -0.1 en valor absoluto.

Fase 4. Construcción del modelo

Para el Modelamiento de los datos se eligieron las técnicas más apropiadas para lograr el objetivo de identificar los factores de riesgo de discapacidad física asociados a la desigualdad social en personas mayores. Una vez realizado un plan de prueba para los modelos escogidos, se procedió a aplicar dichas técnicas sobre los datos para generar los modelos:

1. La división del DataFrame permitió identificar los datos de entrenamiento y los datos de testeo, mediante la identificación de la variable independiente y la variable objetivo.

2. Se ejecutaron los modelos de clasificación: DecisionTreeClassifier, LogisticRegression, neural_network y RandomForestClassifier.

Fase 5. Evaluación del modelo.

Finalmente, se imprimieron las puntuaciones de los cuatro modelos construidos, en donde se evidenció que todos presentan un rendimiento similar, debido a la exactitud que manejan en promedio, con base en lo anterior se puede inferir que cualquiera de los modelos de clasificación implementados, permitiría a las entidades gubernamentales el diseño de estrategias de atención, promoción y prevención de discapacidades físicas en relación a la edad y el estrato socioeconómico de las personas mayores, siendo la regresión logística y la red neuronal, los mejores modelos en cuanto a la exactitud al momento de reconocer la prevalencia de discapacidad en personas mayores:

- DecisionTreeClassifier 80.67%
- Logistic regression 83.56 %
- Red neuronal 83.56 %
- Random forest 82.97 %

Resultados

Las Naciones Unidas (UN) y la OMS, en el marco del cumplimiento de los Objetivos de Desarrollo Sostenible decretaron la Década del Envejecimiento Saludable 2020-2030 cuyo objetivo principal es mejorar la vida de las personas mayores, sus familias y sus comunidades mediante la constitución de sociedades para todas las edades, esto debido el fenómeno de envejecimiento poblacional. Al analizar la evolución de la población por grupos etarios entre el año 1950 y una proyección hasta el año 2070, se advierte la marcada tendencia al decrecimiento que presentarán las franjas etarias relacionadas con los menores de 15 años y los adultos entre los 15 a los 59 años de edad, por su parte las personas mayores presentarán de forma permanente una tendencia al crecimiento.

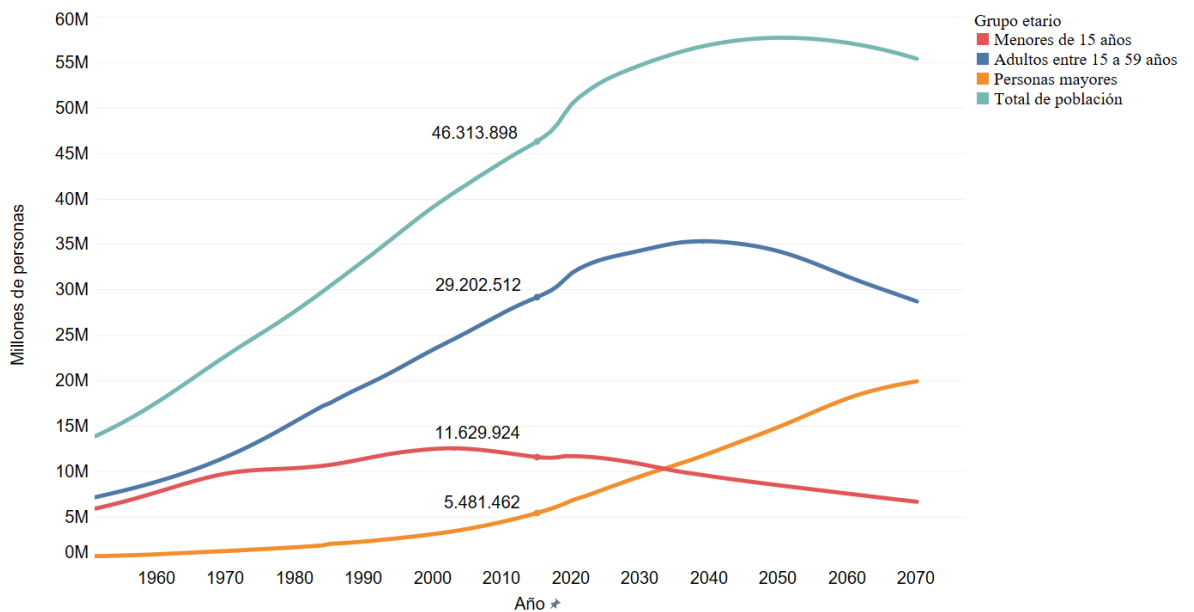


Figura 1. Evolución de la población por grupos etarios, Colombia 1950 – 2070.
Fuente: Elaboración propia a partir de (DANE, 2023)

En cuanto a la composición de la población colombiana para la vigencia 2015, año en el que se adelantó la encuesta SABE, tras calcular la distribución relativa por edad. El 11,84% de la población era mayor de 60 años, y por cada 100 menores de 15 años había 47 personas mayores en el país; para el año 2021

cuando se celebró el XV Congreso Internacional de Envejecimiento y Vejez: “Década del Envejecimiento Saludable 2020 – 2030”, la distribución relativa alcanzó el 13,89%, mientras que las estimaciones para el indicador de envejecimiento evidenciaron que hubo 61 personas mayores por cada 100 menores de 15 años. Según las proyecciones poblacionales, para la finalización de la Década del Envejecimiento Saludable el 17,40% de la población supere los 60 años de edad y por cada 100 menores serán 88 personas mayores, cifras que soportan la avanzada fase de transición demográfica que atravesará la población colombiana y la necesidad de trabajar de forma interdisciplinaria en estrategias de atención a este grupo poblacional.

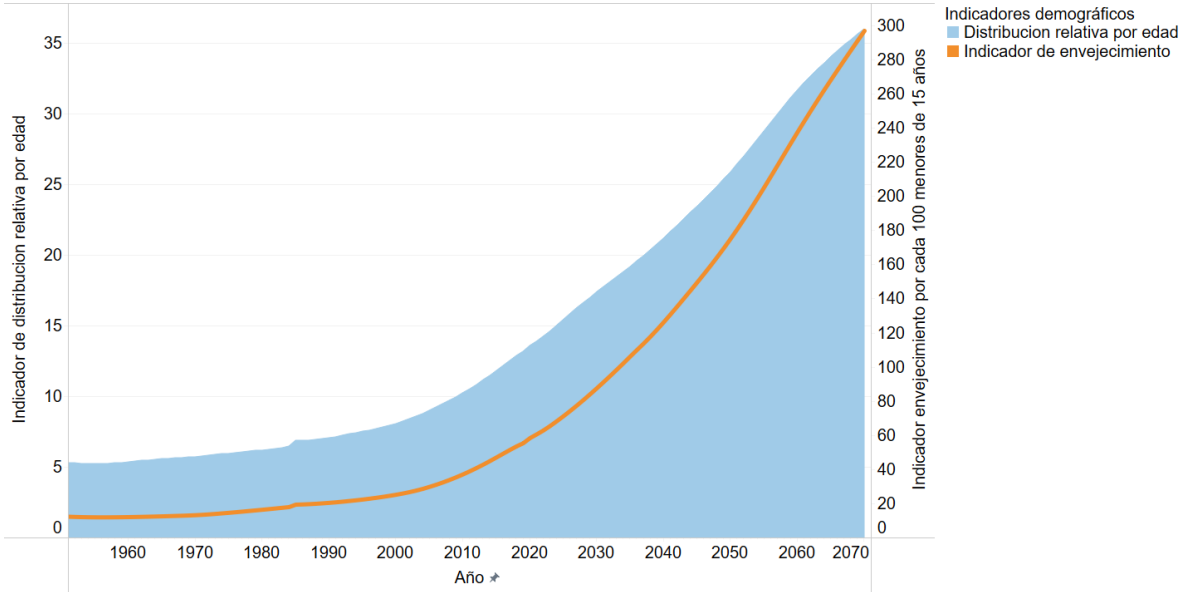


Figura 2. Evolución de la población por grupos etarios, Colombia 1950 – 2070. Fuente: Elaboración propia a partir de (DANE, 2023)

Ante el acelerado proceso de transición demográfica que atraviesa el país, es importante aunar esfuerzos para cumplir con el objetivo trazado en el marco de la Década del Envejecimiento Saludable, que plantea cuatro ámbitos de actuación, de los cuales se destaca, en relación con el tema de interés de la presente investigación, la oferta de atención integrada y centrada en las personas, y los servicios de salud primarios que respondan a las necesidades de las personas

mayores; y proporcionar acceso a la atención a largo plazo para las personas mayores que lo necesiten (Organización Mundial de la Salud, 2022).

En lo tocante a la encuesta SABE Colombia, en ella participaron 23694 personas mayores, con respecto límite arbitrario de edad establecido en países con características de desigualdad muy marcadas, 60 años fue la edad mínima para participar en el estudio, la edad máxima identificada fue de 108 años en el universo de estudio; el promedio de edad de los participantes es de 70.82 años (std 8.2), el 75% de los individuos tenían 76 años o menos. El 57,32% de los encuestados son hombres y el 42,68% mujeres, quienes se encuentran ubicados mayoritariamente en zonas urbanas del país con el 80,93% de los casos.

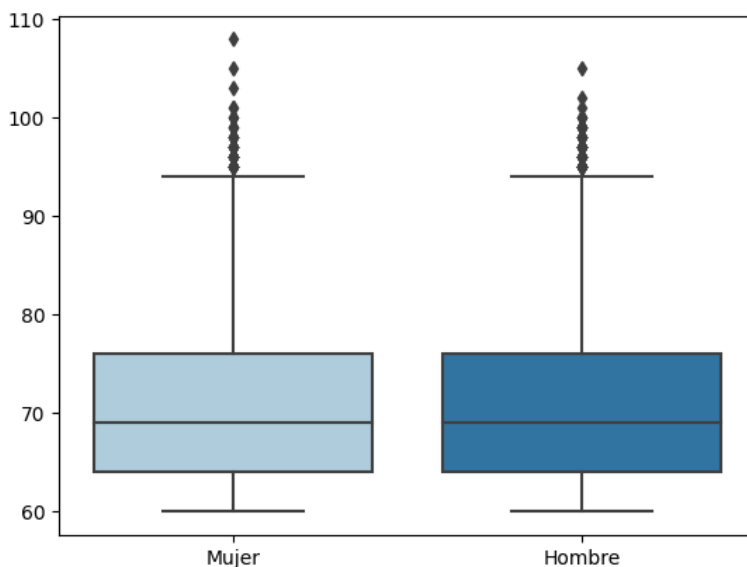


Figura 3. Distribución de edad en relación al sexo
Fuente: Elaboración propia a partir de (SABE, 2016)

Las anteriores cifras, respecto al promedio de la edad de la población, son congruentes con el incremento acelerado en el indicador de envejecimiento y con la esperanza de vida al nacer² que para el quinquenio 2010-2014 fue de 75,22

² Número promedio de años que viviría una persona, siempre y cuando se mantengan las tendencias de mortalidad existentes en un determinado período.

años para el total de la población; las mujeres viven en promedio 6,47 años más que los hombres. Entre los años 2015-2019, la esperanza de vida al nacer al nacer fue de 76,15 años para ambos sexos, la brecha entre hombres y mujeres se redujo a 6,31 años (DANE, 2022).

La pertenencia étnica es uno de los determinantes de las inequidades que enfrentan diferentes grupos poblacionales, fenómeno acentuado en particular en grupos étnicos minoritarios como los raizales, rom o gitanos, afrocolombianos e indígenas. La pertenencia de las personas mayores fue mayoritariamente indígena y afrocolombiana; en lo relacionado a la autodeterminación étnica de los individuos en su mayoría se reconocieron como mestizos y blancos.

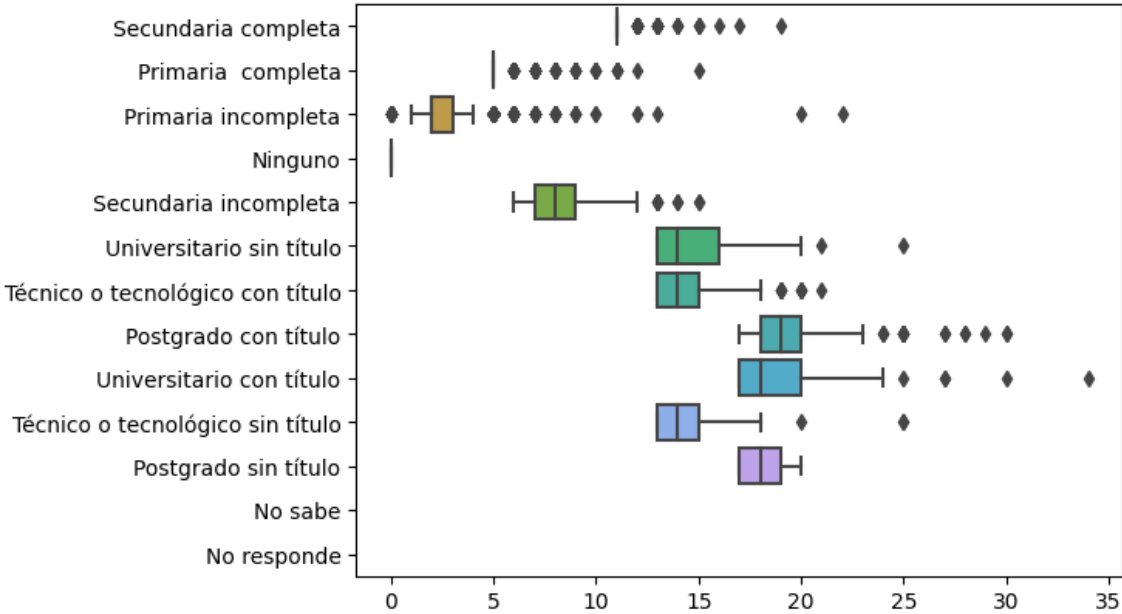


Figura 4. Nivel educativo por años de educación
Fuente: Elaboración propia a partir de (SABE, 2016)

El nivel educativo es fundamental en la estructura socioeconómica de las poblaciones al estar estrechamente ligado con el nivel de ingresos y, por ende, con la desigualdad. En Colombia el bajo logro educativo continúa siendo una problemática latente, en el caso de los adultos mayores, en promedio estudiaron 3,78 años (std 3,91), el 75% de ellos sólo tuvieron hasta 5 años de formación

académica, el 25% restante refleja una gran dispersión de los datos dado que el promedio de años de formación es menor a la desviación estándar, es así como el valor máximo alcanzado durante el curso de vida de algunos individuos es de 34 años. El bajo promedio en los años de educación obedece a que el 40,30% de las personas mayores en Colombia, para la vigencia 2015, no alcanzó a terminar sus estudios de primaria y en promedio estudiaron 2,54 años; el 22,07% no tuvo ningún nivel de educación; respecto al nivel educativo por sexo, el 22,18% de las mujeres no tuvieron ningún nivel de formación académica, mientras que en el caso de los hombres el porcentaje fue de 21,98%, es destacable que, pese a la brecha de género existente, el 2,25% de las mujeres obtuvieran un título universitario, por su parte solo el 1,31% de los hombres lo obtuvo; en cuanto a los procesos de alfabetización, en su mayoría, el 76,50%, de las personas mayores saben leer y escribir, sin embargo, el 21,16% no sabe ni leer ni escribir.

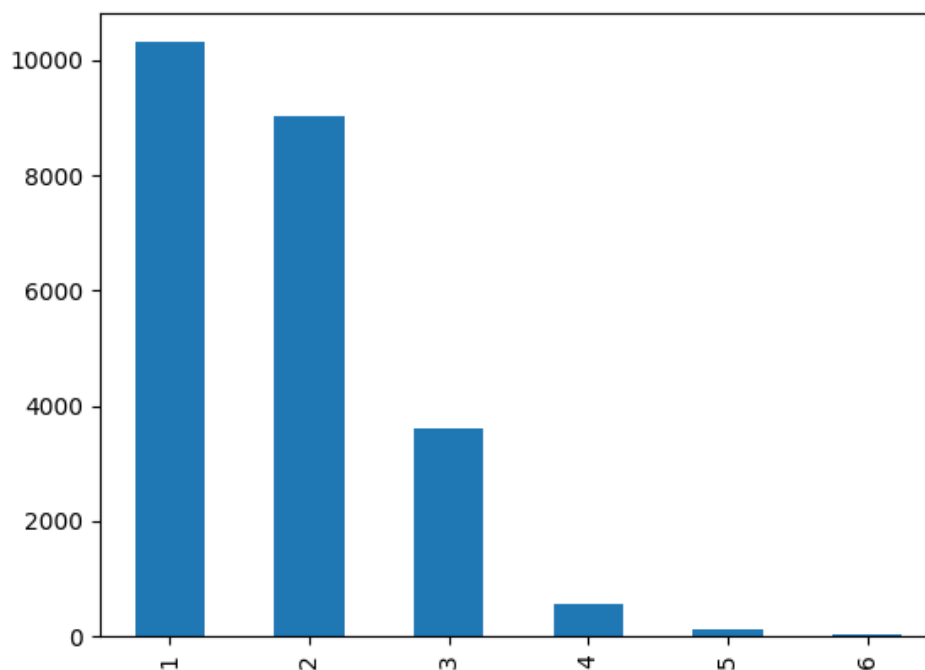


Figura 5. Estrato socioeconómico
Fuente: Elaboración propia a partir de (SABE, 2016)

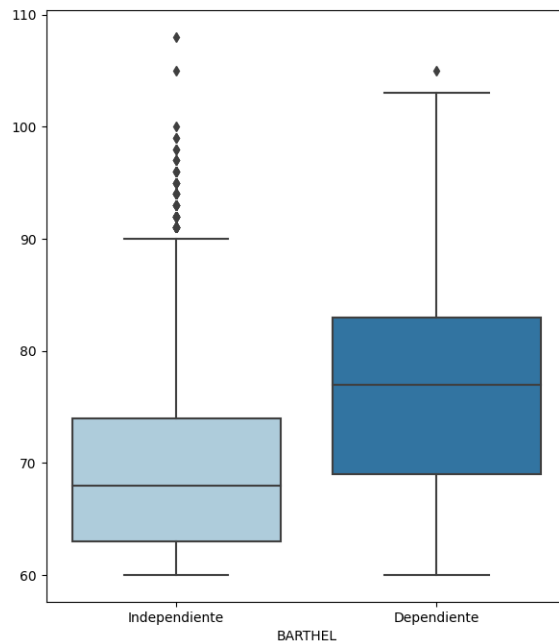
Dentro de las características adquiridas de los individuos se encuentra el estado civil y el estrato socioeconómico. En cuanto a los arreglos familiares, más

de la mitad de las personas mayores en Colombia se encontraban durante la vigencia de estudio casados 53%, viudos el 24,51%, solteros el 11,27 y serados el 11,18%; el 90,71% vivía acompañado y el 9,29 solo. En relación al estrato socioeconómico, pertenece a estratos 1 y 2 catalogados como bajos el 81,65% de las personas mayores; por su parte en cuanto a los niveles 3 y 4, el 17,65% eran estrato medio; finalmente, el 0,7% de las personas mayores de 60 años era de estrato alto. Para el año 2015 en el país el 27,8% de los hogares se encontraban en condición de pobreza monetaria y en términos de desigualdad el índice Gini fue de 0,522 (DANE, 2016), cifras coherentes con el ingreso recurrente de las personas mayores quienes percibían en el 67,48% de los casos menos de un salario mínimo mensual legal vigente (644.350 COP), el 15,87% un salario mínimo y el 9,55% entre uno y dos salarios, y tan solo el 5,32% devengaban más de dos salarios mínimos, el porcentaje restante obedece a quienes no respondieron o no sabían el rango percibido.

Dicho lo anterior alrededor de la situación contextual de las personas mayores en cuanto a su edad, sexo, nivel educativo, arreglos familiares, estrato socioeconómico y nivel de ingresos, es momento de abordar los elementos contemplados alrededor de la capacidad intrínseca de las personas mayores que combina todas las capacidades físicas y mentales de una persona e incluye su capacidad de caminar, pensar, ver, oír y recordar. Distintos factores influyen sobre la capacidad intrínseca como la presencia de enfermedades, los traumatismos y los cambios relacionados con la edad (Organización Mundial de la Salud, 2022).

El índice de Barthel, fue uno de los primeros intentos de cuantificar la discapacidad física, se fundamenta en bases empíricas relacionadas con 10 actividades de la vida diaria (AVD) que pueden realizar, en este caso las personas mayores, entre las que se encuentran alimentación, trasladarse, aseo personal, uso del inodoro, bañarse, desplazarse, subir y bajar escaleras, vestirse y desvestirse, continencia fecal y urinaria, cada una de ellas recibe una puntuación entre 0 y 100 como indicador de independencia; la discapacidad se definió como

la dificultad de una actividad de (Cid-Ruzafa & Este índice utilizado en el área para valorar el discapacidad física vida de las de servicios herramienta de



realizar al menos las mencionadas Moreno, 1997) al ser ampliamente de la salud pública impacto de la en la calidad de personas y el uso sanitarios, es una monitoreo que

debe considerarse en la construcción de políticas públicas coherentes con las necesidades básicas de las poblaciones y con el diseño de estrategias de atención, promoción y prevención de la discapacidad física, esto debido a que a medida que incrementa la edad, incrementa la prevalencia de discapacidad física, lo cual implica que las personas mayores discapacitadas deben contar con red de apoyo manifestado a través del número de convivientes el número de convivientes; también es importante resaltar que la prevalencia de la discapacidad física se podría relacionar con los años de educación presentado por las personas, debido a que a mayor nivel educativo, menor prevalencia de discapacidad.

Figura 6. Distribución de edad en relación al índice de Barthel

Fuente: Elaboración propia a partir de (SABE, 2016)

El 22,07% de las personas mayores en Colombia para el año 2015 se encontraban en condición de discapacidad con base en el índice de Barthel; al analizar la correlación del índice con las características adscritas al individuo como la edad, se puede advertir una correlación moderada, directamente proporcional, entre la edad avanzada y la prevalencia de la discapacidad, a mayor edad, mayor nivel de dependencia, estos elementos determinantes en la construcción de los modelos predictivos. La dependencia para el desarrollo de las actividades de la vida diaria también se relaciona con la necesidad de soporte por parte de la red de apoyo familiar de las personas mayores, por consiguiente, se incrementa el número de convivientes a medida que incrementa la edad.

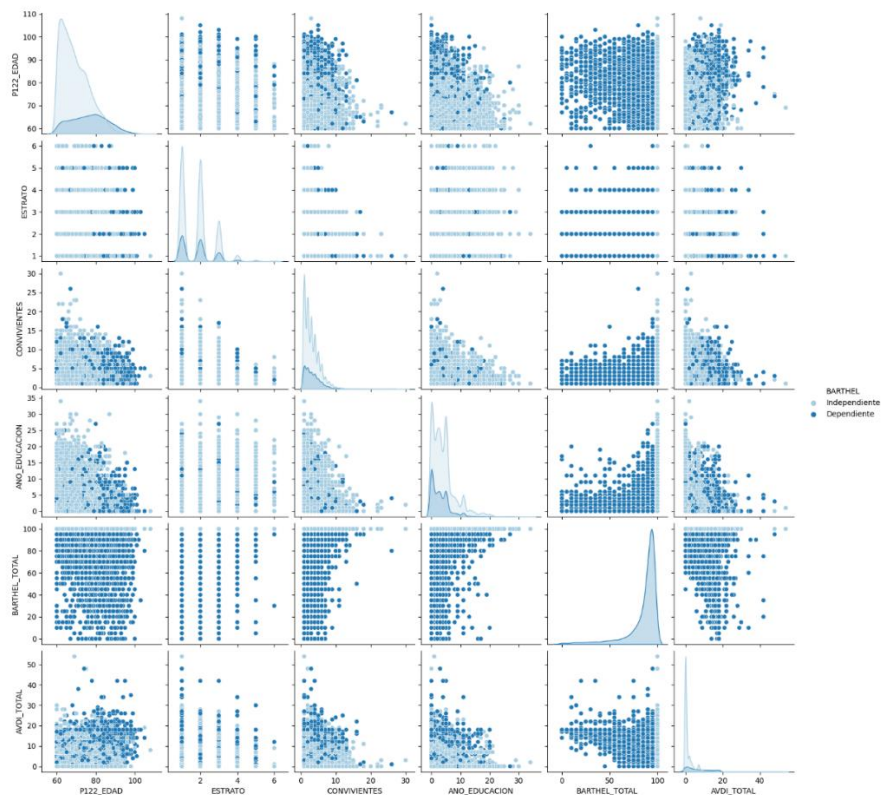


Figura 7. Correlación del Índice de Barthel
 Fuente: Elaboración propia a partir de (SABE, 2016)

Discusión

Con fundamento en la situación contextual alrededor del acelerado proceso de envejecimiento que atraviesa la población colombiana, aunado al contexto de desigualdad en el que envejecen y la prevalencia de la discapacidad física en las personas mayores, se decidió recurrir a la ciencia de datos, específicamente a los modelos de clasificación cuyo algoritmo permite la asignación de etiquetas o categorías a un conjunto de datos para clasificarlos, esta técnica es fundamental en el aprendizaje automático y en la minería de datos; la aplicación de los modelos de clasificación implican un proceso de entrenamiento del modelo con la utilización de un subconjunto de datos, la intención es lograr el aprendizaje del modelo para identificar los patrones y relaciones en los datos, para luego aplicar dicho

conocimiento en el subconjunto restante de datos que no tienen etiquetas y lograr predecir la clase a la que estos pertenecen (Cady, 2017).

La elección del modelo de clasificación depende del tipo de datos, la naturaleza del problema y otros factores, en este estudio se decidió recurrir a diversos modelos de clasificación como el árbol de decisión, la regresión logística, la red neuronal artificial y el bosque aleatorio para lograr predecir la discapacidad física medida por el índice de Barthel en relación a condiciones adscritas y adquiridas de los individuos. A continuación, se describe el proceso de construcción de los modelos y los resultados obtenidos:

Se importan las librerías necesarias para el procesamiento de los datos

1. Se configuró el entorno de ejecución en línea de Google basado en Jupyter Notebook para conectar Drive con el entorno colaborativo *Colab*.
2. Se realizó la función `pd.read_excel` para leer datos de un archivo Excel y cargarlos en un DataFrame; posteriormente este se visualizaron las 4 primeras filas que lo componen y las 34 columnas que lo componen.
3. Se utilizó el atributo `columns` que permitiera conocer los nombres de las variables que componen el DataFrame.
4. Posteriormente se procedió a identificar en un mapa de calor los valores nulos en el DataFrame para decidir si es procedente eliminar las variables que presentan un alto número de datos perdidos o si se debe sólo eliminar los valores perdidos en mención.

Este mapa de calor evidencia las falencias alrededor del registro de información que advierte las características socioeconómicas de las personas mayores en cuanto a su situación laboral, los ingresos percibidos y la percepción de satisfacción con los mismos. De igual forma, en lo referente a los procesos identitarios de reconocimiento étnico, se presenta gran concentración de valores nulos.

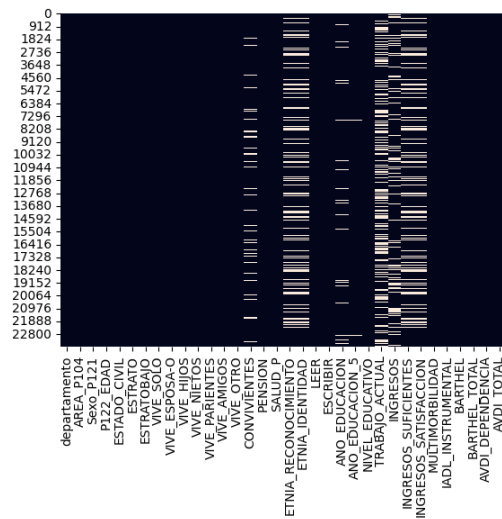


Figura 8. Mapa de calor de valores nulos

Fuente: Elaboración propia a partir de (SABE, 2016).

5. Se creó un segundo DataFrame con valores nulos para identificar su frecuencia.
6. Se procedió a eliminar las columnas que concentraban gran cantidad de datos perdidos, dado que inciden en la calidad de los datos. Se decidió no implementar el método dropna() que eliminaría las filas que contienen datos nulos y podría reducir de forma drástica el DataFrame. Para iniciar propiamente el análisis exploratorio se obtuvo información sobre el tipo de cada una de las variables que conforman la base de datos y de los valores no null.

Se identificaron errores en cuanto al tipo de variable de ESTRATO como entero (int), CONVIVIENTES y ANO_EDUCACION como decimales (float).

7. Adicional a la depuración de las anteriores columnas, se procedió a eliminar información relacionada con los convivientes al ser variables dummies que no aportan mucha información en el modelo.
8. Se procedió con el cambio de tipo de variables, ESTRATO pasó a ser categórica (object), CONVIVIENTES y ANO_EDUCACION como enteros (int).
9. Una vez modificados los tipos de variables se realizó el resumen estadístico de las variables int para dar inicio a una descripción de las características sociodemográficas de las personas mayores en Colombia (ver Resultados), con el objetivo de conocer la situación contextual de quienes participaron en el estudio y de esta forma se realizó una aproximación a los posibles modelos a implementar y las variables a considerar en el modelamiento.

10. Se inició el análisis descriptivo de los datos para identificar el comportamiento del fenómeno y las variables para someter a estudio.
11. En la construcción del árbol de decisión se cargaron de forma inicial las variables 'departamento', 'AREA_P104', 'Sexo_P121', 'P122_EDAD', 'ESTADO_CIVIL', 'ESTRATO', 'ESTRATODBAJO', 'VIVE_SOLO', 'CONVIVIENTES', 'PENSION', 'SALUD_P', 'ETNIA_RECONOCIMIENTO', 'LEER', 'ESCRIBIR', 'ANO_EDUCACION', 'ANO_EDUCACION_5', 'NIVEL_EDUCATIVO', 'INGRESOS', 'MULTIMORBILIDAD', 'IADL_INSTRUMENTAL', 'BARTHEL', 'BARTHEL_TOTAL', 'AVDI_DEPENDENCIA', 'AVDI_TOTAL'.
12. Debido a que el modelo no podía realizarse con todas estas variables categóricas, se procedió a eliminar información mediante la función drop relacionada con las variables departamento', 'ESTRATODBAJO', 'VIVE_SOLO', 'CONVIVIENTES', 'PENSION', 'SALUD_P', 'ETNIA_RECONOCIMIENTO', 'LEER', 'ESCRIBIR', 'ANO_EDUCACION', 'NIVEL_EDUCATIVO', 'INGRESOS', 'MULTIMORBILIDAD', 'IADL_INSTRUMENTAL', 'BARTHEL_TOTAL', 'AVDI_DEPENDENCIA', 'AVDI_TOTAL'.
13. Al depurar las variables se realizó la transformación de estas categóricas en dummies mediante la función pd.get_dummies.
14. Se identificaron las variables dummies.loc como predictoras (X) y BARTHEL como la variable a predecir (Y).
15. Se procedió con la configuración del modelo de entrenamiento y el modelo de prueba.
16. Como resultado de la aplicación del tree.DecisionTreeClassifier se obtuvo árbol_discapacidad:

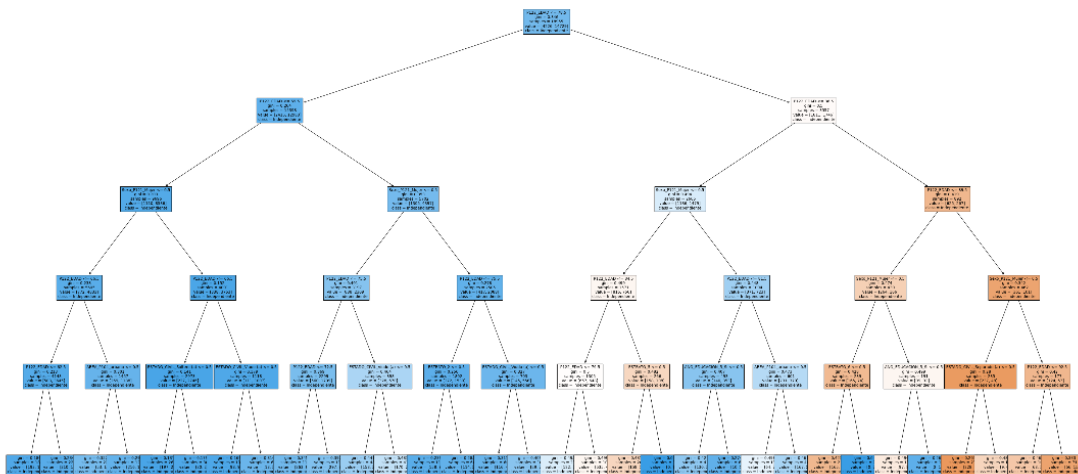


Figura 9. Modelo de Árbol de decisión
Fuente: Elaboración propia a partir de (SABE, 2016).

17. Este árbol permite la clasificación de la discapacidad en relación a la edad, el sexo, estado civil, el nivel educativo, la matriz de confusión, como herramienta de evaluación del rendimiento del modelo de clasificación, esta matriz muestra la cantidad de observaciones que fueron clasificadas correctamente en cada clase y las que fueron clasificadas de forma incorrecta, las clases se refieren a los verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN).

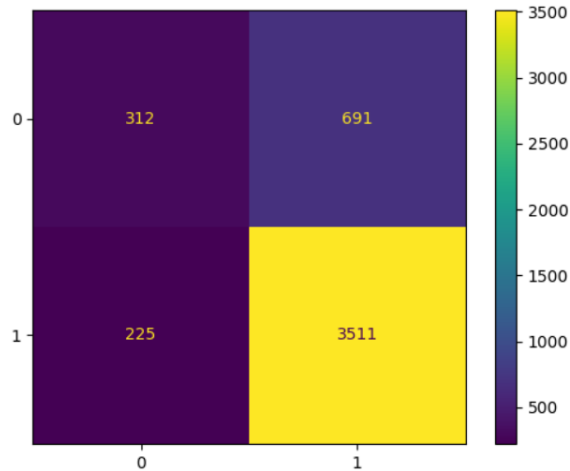


Figura 10. Matriz de confusión del modelo de Árbol de Decisión
Fuente: Elaboración propia a partir de (SABE, 2016).

18. La exactitud del modelo mide la proporción de las observaciones clasificadas correctamente y se calcula como $(TP+TN) / (TP+TN+FP+FN)$, el resultado obtenido indicó que el árbol de decisión construido alcanzó una exactitud del 80,67%.

Los resultados de este estudio evidencian la pertinencia de diseñar un modelo predictivo para la identificación de factores de desigualdad social asociados a la discapacidad en personas mayores en el marco de los determinantes de envejecimiento saludable, todo esto analizado desde un enfoque multidimensional que conjugó la ciencia de datos para la construcción de un algoritmo que lograra generar sistemas de identificación de las personas mayores potencialmente vulnerables de sufrir de discapacidad física durante la vejez como consecuencia de determinantes estructurales relacionados con características sociodemográficas como la edad avanzada. Finalmente, esta aproximación analítica sobre la pertinencia de los modelos predictivos es acorde con la Década

del Envejecimiento Saludable al ser una apuesta organizacional que convoca a los Estados a diseñar estrategia que impacten de forma positiva el contexto de desigualdad de los individuos, para lograr la reducción de las inequidades estructurales que se prolongan con el pasar de los años.

Referencias

- DuMontier, C., Clough-Gorr, K., Silliman, R., Atascado, A., & Moser, A. (2018). Health-Related Quality of Life in a Predictive Model for Mortality in Older Breast Cancer Survivors. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 1115-1122.
- Brown, T., & Hargrove, T. (2018). Psychosocial Mechanisms Underlying Older Black Men's Health. *JOURNALS OF GERONTOLOGY SERIES B-PSYCHOLOGICAL SCIENCES AND SOCIAL SCIENCES*, 188-197.
- Brundle, C., Heaven, A., Brown, L., Teale, E., Young, J., West, R., & Clegg, A. (2019). Convergent validity of the electronic frailty index. *AGE AND AGEING*, 152-156.
- Cady, F. (2017). *The Data Science Handbook*. Wiley.
- Cárdenas, C., González, S., Nahuel, R., Herrera, P., Ferrada, L., & Celis, D. (2018). Design of a predictive model of cardiovascular screening using decision trees: propensity of patients to present type 2 diabetes, arterial hypertension or dyslipidemia. Pilot study commune of Quellón, Chiloé. *Revista chilena de cardiología*.
- Cid-Ruzafa, J., & Moreno, J. D. (1997). Valoración de la discapacidad física: el índice de Barthel. *Revista Española de Salud Pública*, 71(2).
- Colombia, P. d. (2 de mayo de 2022). Por medio del cual se adiciona el Capítulo 7 al Título 2 de la Parte 9 del Libro 2 del Decreto 780 de 2016 relativo a la Política Pública Nacional de Envejecimiento y Vejez 2022 - 2031.
- DANE. (Marzo de 2016). *Boletín técnico pobreza monetaria y multidimensional en Colombia 2015*. Obtenido de DANE Información para todos: https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/bol_pobreza_15_.pdf
- DANE. (2021). *Medida de Pobreza Multidimensional Municipal de Fuente Censal*. Bogotá D.C.
- DANE. (2022). *DANE Información para todos*. Obtenido de Indicadores demográficos 1985 - 2020: [http://www.dane.gov.co/index.php?option=com_content&view=article&id=853&Itemid=28&phpMyAdmin#:~:text=La%20esperanza%20de%20vida%20\(que,a%C3%B1os%20m%C3%A1s%20que%20los%20hombres](http://www.dane.gov.co/index.php?option=com_content&view=article&id=853&Itemid=28&phpMyAdmin#:~:text=La%20esperanza%20de%20vida%20(que,a%C3%B1os%20m%C3%A1s%20que%20los%20hombres).
- DANE. (22 de Marzo de 2023). *Proyecciones y retroproyecciones de población nacional para el periodo 1950-2019 y 2020-2070 con base en el CNPV 2018*. Obtenido de DANE Información para todos: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>

- de Lange, D. W., Brinkman, S., Flaatten, H., Boumendil, A., Morandi, A., Andersen, F., . . . Christensen, S. (2019). Cumulative Prognostic Score Predicting Mortality in Patients Older Than 80 Years Admitted to the ICU. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 1263-1267.
- Dirección de Epidemiología y Demografía. (2018). *Documento Metodológico Encuesta Nacional de Salud, Bienestar y Envejecimiento SABE Colombia*. Bogotá D.C.
- Earnshaw, R., Dill, J., & Kasik, D. (2019). *Data Science and Visual Computing*. Springer. doi:10.1007/978-3-030-24367-8_1
- Ek, S., Rizzuto, D., Fratiglioni, L., Calderón-Larrañaga, A., Johnell, K., Sjoberg, L., . . . Welmer, A.-K. (2019). Risk Factors for Injurious Falls in Older Adults: The Role of Sex and Length of Follow-Up. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 246-253.
- Ferraro, K. F., & Shippee, T. P. (2009). Aging and Cumulative Inequality: How Does Inequality Get Under the Skin? *The Gerontologist*, 49(3), 333-343.
- Folstein, M., Folstein, S., & McHugh, P. (Nov de 1975). A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. doi:10.1016/0022-3956(75)90026-6
- Gale, C. R., Westbury, L., Cooper, C., & Dennison, E. (2018). Risk factors for incident falls in older men and women: the English longitudinal study of ageing. *BMC GERIATRICS*.
- Gómez, F., Osorio García, D., Panesso, L., & Curcio, C. L. (2021). Healthy aging determinants and disability among older adults: SABE Colombia. *Pan American Journal of Public Health*. doi:10.26633/RPSP.2021.98
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Nueva York: Springer. doi:10.1007/978-0-387-84858-7_2
- How to Develop Data Science Skills. (2015). En C. Shan, H. Wang, W. Chen, & M. Song, *The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists* (págs. 115 - 122).
- Icaza, M., & Albala, C. (1999). Minimental State Examinations (MMSE) of the madness study in Chile : statistical analysis. *Pan American Health Organization*.
- Joyanes Aguilar, L. (2017). *Industria 4.0 La Cuarta Revolución Industrial*. México: Algaomega Grupo Editor, S.A.

- Lai, J., Wang, H., Pan, Z., & Su, F. (2019). A novel six-microRNA-based model to improve prognosis prediction of breast cancer. *AGING-US*, 649-662.
- Lisk, R., Uddin, M., Parbhoo, A., Yeong, K., Fluck, D., Sharma, P., . . . Han, T. (2019). Predictive model of length of stay in hospital among older patients. *AGING CLINICAL AND EXPERIMENTAL RESEARCH*, 993-999.
- Mamoshina, P., Kochetov, K., Putin, E., Cortese, F., Aliper, A., Lee, W.-S., . . . Kovalchuk, O. (2018). Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. *JOURNALS OF GERONTOLOGY SERIES A-BIOLOGICAL SCIENCES AND MEDICAL SCIENCES*, 1482-1490.
- Ministerio de Salud y Protección Social - Departamento Administrativo de Ciencia Tecnología e Innovación, COLCIENCIAS, Universidad del Valle y Universidad de Caldas. (2016). *Encuesta SABE Colombia: Vejez y Calidad de vida en Colombia*. Bogotá D.C.: PREGRAF IMPRESORES S.A.S.
- Musich, S., Wang, S., Ruiz, J., Hawkins, K., & Wicker, E. (2018). The impact of mobility limitations on health outcomes among older adults. *GERIATRIC NURSING*, 162-169.
- Organización Mundial de la Salud. (2015). *Informe mundial sobre el envejecimiento y la salud*. Organización Mundial de la Salud.
- Organización Mundial de la Salud. (diciembre de 2022). *Decade of Healthy Ageing*. Obtenido de Organización Mundial de la Salud: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020rev-es.pdf?sfvrsn=b4b75ebc_25&download=true
- Park, S.-H. (2018). Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. *AGING CLINICAL AND EXPERIMENTAL RESEARCH*, 1-16.
- Piatetsky-Shapiro, G. (2000). Knowledge Discovery in Databases: 10 years after. *SIGKDD Explorations*, 61.
- Provost, F., & Fawcett, T. (Marzo de 2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1). doi:10.1089/big.2013.1508
- Rosso, A., Metti, A., Faulkner, K., Brach, J., Studenski, S., Redfern, M., & Rosano, C. (2019). Associations of Usual Pace and Complex Task Gait Speeds With Incident Mobility Disability. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 2072-2076.
- Schröera, C., Kruse, F., & Jorge, M. G. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 526-534.

- Sen, A. K. (1999). *Development as Freedom*. Nueva York: Oxford University Press.
- Shi, S., McCarthy, E., Mitchell, S., & Kim, D. (2020). Changes in Predictive Performance of a Frailty Index with Availability of Clinical Domains. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 1771-1777.
- Therborn, G. (2015). *La Desigualdad Mata*. Madrid: Alianza Editorial.
- Wang, D., Yao, J., Zirek, Y., Reijnierse, E., & Maier, A. (2020). Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *JOURNAL OF CACHEXIA SARCOPENIA AND MUSCLE*, 3-25.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining. Practical Machine Learning Tools and Techniques* (Third ed.). United States: Morgan Kaufmann Publishers.
- Womens Hlth Initiative Investigato. (2018). 36-Item Short Form Survey (SF-36) Versus Gait Speed As Predictor of Preclinical Mobility Disability in Older Women: The Women's Health Initiative. *JOURNAL OF THE AMERICAN GERIATRICS SOCIETY*, 66, 706-713.
- Zeitzer, J., Blackwell, T., Hoffman, A., Cummings, S., Ancoli-Israel, S., & Piedra, K. (2018). Daily Patterns of Accelerometer Activity Predict Changes in Sleep, Cognition, and Mortality in Older Men. *JOURNALS OF GERONTOLOGY SERIES A-BIOLOGICAL SCIENCES AND MEDICAL SCIENCES*, 682-687.
- Zhang, H., Zhu, Y., Hao, M., Wang, J., Wang, Z., Chu, X., . . . Wang, X. (2021). The Modified Healthy Ageing Index Is Associated with Mortality and Disability: The Rugao Longevity and Ageing Study. *GERONTOLOGY*.
- Zunzunegui, M. V., Alvarado, B. E., Guerra, R., Gómez, J. F., Ylli, A., & Guralnik, J. M. (2015). The mobility gap between older men and women: the embodiment of gender. *Arch Gerontol Geriatr*, 140-148.