INVESTIGACIÓN e INNOVACIÓN en INGENIERÍAS
ISSN2344-8652

Identificación de las tendencias del estado actual de la extraedad en el acceso a la educación básica en Colombia aplicando inteligencia artificial

Identification of trends in the current state of extra age in access to basic education in Colombia applying artificial intelligence

Hugo Ordoñez **Cristian Ordoñez**
Universidad del Cauca, Colombia

Victor Buchelis
Universidad del Valle, Colombia

OPEN ACCESS

Recibido: 13/04/2022
Aceptado: 04/07/2022
Publicado: 07/10/2022

Correspondencia de autores:
hugoordonez@unicauca.edu.co

Resumen

Objetivo: Proponer un modelo de predicción basado en inteligencia artificial para predecir la tendencia de extraedad en la matricula el sistema educativo colombiano. **Metodología:** El sistema educativo colombiano ha experimentado un cambio fundamental de transformación. En consecuencia, la ampliación de la matrícula refleja también cambios en la composición de la población según edad, donde se tiene que los rangos tentativos de edad poco comunes para acceder a la educación. Por ejemplo, un estudiante de segundo grado debe tener entre 7 y 8 años de edad, si tiene entre 9 o más años, es un estudiante en extraedad. El escenario de extraedad está vinculado a dos aspectos: 1), ingreso tardío al sistema educativo. 2) Altas tasas de repitencia. En este sentido, la extraedad se convierte en una dificultad de contexto favorable para la exclusión escolar. **Resultados:** El modelo fue evaluado con las métricas coeficiente de determinación (R2), raíz del error cuadrático medio (RMSE) y error porcentual absoluto medio (MAPE). Los resultados obtenidos demuestran que el modelo puede servir de base para la toma de decisiones a entes gubernamentales para la implementación de políticas orientadas estrategias pedagógicas, estrategias didácticas, en miras de la diminución de la extraedad. **Conclusiones:** El modelo propuesto, pude servir de base en el ministerio de educación para la implementación de políticas de orientación administrativas para la atención educativa de la población en extraedad, joven, adulta y adulta mayor con discapacidad intelectual y psicosocial, en el marco de la inclusión y la equidad en la educación tanto rural como urbana.

Palabras clave: Extraedad, educación, edad, predicción, machine learning.

Abstract

Objective: Propose a prediction model based on artificial intelligence to predict the trend of overage enrollment in the Colombian educational system. **Methodology:** The Colombian educational system has undergone a fundamental change of transformation. Consequently, the increase in enrollment also reflects changes in the composition of the population according to age, where the tentative age ranges are uncommon to access education. For example, a second grade student must be between 7 and 8 years old, if they are between 9 and older, they are an overage student. The overage scenario is linked to two aspects: 1), late entry into the educational system. 2) High repetition rates. In this sense, being over-aged becomes a favorable context difficulty for school exclusion. **Results:** The model was evaluated with the coefficient of determination (R2), root mean square error (RMSE) and mean absolute percentage error (MAPE) metrics. The results obtained show that the model can serve as a basis for decision-making to government entities for the implementation of policies oriented pedagogical strategies, didactic strategies, with a view to reducing overage. **Conclusions:** The proposed model could serve as a basis in the Ministry of Education for the implementation of administrative orientation policies for the educational attention of the population in overage, young, adult and older adults with intellectual and psychosocial disabilities, within the framework of the inclusion and equity in both rural and urban education.

Keywords: Age, extra age, education, prediction, machine learning.

Como citar (IEEE): H. Ordoñez., C .Ordoñez., y V. Buchelis. "Identificación de las tendencias del estado actual de la extraedad en el acceso a la educación básica en Colombia aplicando inteligencia artificial". Investigación e Innovación en Ingenierías, vol. 10, n°2, 106-118, 2022. DOI: https://doi.org/10.17081/invinno.10.2.6130

Introducción

La educación debe ser vista como un derecho fundamental de toda persona, y debe ser garantizado a lo largo de su vida. Sin embargo, el acceso a la educación debe ir acompañado de calidad como un pilar del desarrollo nacional. Esta es la visión establecida por las Naciones Unidas en 2015 cuando promulgó el Desarrollo Sostenible (ODS) [1], como el objetivo para alcanzar un futuro mejor y sostenible para todos [2 ,3]. La educación promueve el cambio socioeconómico ascendente y es un factor fundamental para disminuir los índices de pobreza. En los últimos años, según [4], se aumentó considerablemente el acceso a la educación, de tal forma que las tasas de matriculación en las escuelas aumentaron en todos los niveles.

Considerando la educación como eje primordial , para [5] el sistema educativo colombiano ha experimentado un cambio fundamental de transformación, obteniendo como el resultado más visible la expansión en el acceso en todos los niveles de formación, a través de políticas ambiciosas para eliminar las barreras en la matriculación [6] , buscando así llevar los servicios educativos a todos los rincones del país. Es así, como la matrícula ha aumentado considerablemente en todos los niveles. En solo una década, la participación en la educación Preescolar, Básica, Media se ha duplicado, hasta el 40 % y el 50 %, respectivamente [7].

La diversificación de la matrícula muestra los cambios en la estructura de la población según edad [8], en este sentido, en Colombia se tiene que los rangos tentativos de edad promedio para acceder a la educación son: Transición (5 años), Primaria (6 a 10), Secundaria (11 a 14), Media (15 a 16) y Superior (17 a 21). Según la Ley General de Educación se ha definido que la educación es necesaria entre los 5 y 15 años de edad, en los grados de transición a noveno, y de la misma forma el grado de preescolar obligatorio (transición) lo realizan los niños entre 5 y 6 años de edad. Por ejemplo, un estudiante de segundo grado debe estar en el rango de edad de 7 y 8 años, si tiene entre 9 o más años, es un estudiante en extraedad. Entonces, la extraedad es el desnivel entre la edad y el grado en curso, que sucede cuando un niño, joven o adulto tiene un cierto número de años más, sobre el nivel de la edad promedio deseada para cursar un grado en específico[8].

La situación de extraedad está asociada a dos aspectos: 1), la entrada tardía al sistema educativo, lo cual es visto de manera deficiente para América Latina. 2) Grandes cifras de repitencia. La repitencia, cuando se combina con el inicio tardío del primer grado, crea un contexto propicio para la exclusión escolar [9] [10], lo que se considera como una decepción escolar en sectores desvalidos de la sociedad [11]. En consecuencia, se hace necesario diseñar estrategias que aporten a la toma de decisiones nacionales y locales para hacer realidad la matriculación universal. Esta expansión debe priorizarse donde los niveles de inscripción al sistema educativo son los más bajos.

Como apoyo al diseño de estrategias para la toma decisiones aparecen los algoritmos de aprendizaje automático o de maquina (machine learning) los cuales son parte de la inteligencia artificial[12,13]. Estos algoritmos permiten analizar grandes volúmenes de datos provenientes de ambientes big-data, con el propósito de definir modelos que permitan realizar predicciones. Los modelos de predicción se han utilizado durante mucho tiempo como soporte a la toma de decisiones en muchos dominios, como por ejemplo, en el sector salud para acompañar prácticas clínicas por medio de herramientas destinadas a ayudar a los médicos a definir flujos de trabajo, y así mejorar los resultados en los tratamientos de los pacientes [14]. Para analizar las tendencias de hurto en Colombia [15], en el campo especifico de la educación, para maximización de la educación sostenible [16], para la mejora de la educación superior y adelanto económico regional en Europa [17].

Con base a lo anterior, en este artículo se propone un modelo de predicción basado en inteligencia artificial que implementa algoritmos de machine learning para predecir la tendencia de extraedad en la matricula el sistema educativo colombiano de educación preescolar, de educación básica (primaria cinco grados y secundaria cuatro grados), de educación media (dos grados y culmina con el título de bachiller). El modelo implementa cuatro (regresión lineal, Arboles de decisión, regresión Lasso y regresión Rigde) algoritmos de machine learning que permiten hacer predicciones a través de regresión. El modelo utiliza un dataset con 6916617 registros y 34 columnas. El modelo se plantea con el fin de servir de base en la toma de decisiones en entidades territoriales certificadas (departamentos, municipios y distritos), entes educativos, interesadas en la atención de niños, niñas y jóvenes en extraedad escolar, para la implementación de modelos aceleración del aprendizaje o estrategias de facilitación de acceso a la educación.

El presente artículo está organizado de la siguiente manera, la sección dos presenta el escenario de motivación, la sección tres describe el modelo propuesto, la sección cuatro los resultados de evaluación y finalmente la sección cinco las conclusiones y el trabajo a futuro.

Metodología

En Colombia, el gobierno nacional ha implementado políticas para que los niños, las niñas y los adolescentes accedan a una educación de calidad, la cual tiene carácter obligatorio por parte del estado en los años cubiertos entre el preescolar y noveno de educación básica [5] . Según el Ministerio de Educación Nacional (MEN), los retos en el sistema educativo son cada vez más complejos y se relacionan con el acceso a los programas de educación oportunos y de calidad. El MEN, informa que en Colombia las niñas, niños y adolescentes no llevan a cabo sus procesos educativos de forma acertada y cumplida. A pesar de que se ha incrementado la cobertura (96,4% en 2017). La educación en Colombia presenta el problema de la interrupción de los estudios por parte de los estudiantes, debido a inconvenientes de salud, desplazamiento, trabajo, económicos, mudanza. Esto problemas pueden ser interpretados como causas diferentes de entorpecimiento en el trascurso de formación, los cuales pueden llegar a ser causas comunes, debido a que problemas económicos pueden hacer que alumnos tengan que trabajar para mejorar la mala situación económica de sus hogares, por otra parte, la mudanza y el desplazamiento, se convierten en parte de las mismas causas.

Como base en lo anterior, la Tabla 1, muestra las estadísticas de la extraedad en Colombia. En estas se tiene que para el grado de Transición en la zona rural el 5,08% de los estudiantes se encuentra en extraedad, a pesar de que el acceso a la educación en el área rural es un poco limitado la cifra de extraedad es baja. Para la zona urbana en este mismo grado el 6,47% de los estudiantes está en extraedad, al contrario de la zona rural, la zona urbana tiene mayor oferta educativa ya que en esta zona se encuentra mayor cantidad de establecimientos educativos. En este sentido, el número de estudiantes en extraedad es relativamente alto, esto puede presentarse debido a inconvenientes académicos, sino también de convivencia, matoneo y factores económicos por parte de los padres, a la alta demanda de cupos escolares en el sector oficial.

Para el caso de la primaria, se tiene que en la zona rural el 10,52% de los alumnos se encuentra en extraedad, este número duplica el de la transición, esto se debe a que el sector rural colombiano, presenta el aislamiento, asimismo no hay suficientes salones de clase, a esto se suma el trabajo de los niños quienes buscan colaborar en el ingreso económico de la familia, además la poca escolaridad de los padres, factor que genera señal negativa en el acceso de los niños a la educación. En el caso de la zona urbana para primaria se tiene que el 6,60% de los alumnos se encuentran en extraedad. Es clave que, en algunas ciudades colombianas, la escuela puede estar a una distancia considerable para los niños; Además de las distancias, se presentan factores de orden socioeconómico, cultural y de inseguridad.

En relación a la Secundaria y Media , para la zona rural se tiene que él (43,45 para secundaria y 124,09% para media) de los alumnos se encuentran en extraedad, como se puede apreciar es muy notable el porcentaje, como se mencionó anteriormente en la zona rural el trabajo de los alumnos es fundamental para su familia, a esto se suma que la educación se lleva a cabo en zonas dispersas, apartadas de las poblaciones con mayor explanación, con mínima inversión en infraestructura y tecnología y carente de perfeccionamiento de varias competencias, entre ellas las tecnológicas debido a la ausencia de conectividad. En el caso de la educación media los estudiantes que pasan los 18 años de edad, deciden enfocarse en un trabajo de manera permanente, ya que en muchas ocasiones es el único sustento de su familia. De la misma forma para la zona urbana se tiene (33,59% para secundaria y 115,69% para media), como se puede observar los porcentajes de alumnos en estas etapas de educción son bastante altas, Este escenario se presenta por la obligación de llevar a cabo tareas y trabajos para la manutención familiar, además de la carencia de oferta educativa de calidad en estos niveles, y falta de docentes.

Tabla 1. Estadísticas de extraedad en Colombia según la zona, el sector y el género

Grado	Zona	Sector	Genero	Toral edad ideal	Total extraedad
Transición	Rural	OFICIAL	Masculino	121097	6655
	Rural	NO_OFICIAL	Masculino	2001	219
	Rural	OFICIAL	Femenino	115488	5175
	Rural	NO_OFICIAL	Femenino	1927	172
	Urbana	OFICIAL	Masculino	64617	4350
	Urbana	NO_OFICIAL	Masculino	48990	3819
	Urbana	OFICIAL	Femenino	64176	3513
	Urbana	NO_OFICIAL	Femenino	47695	2908
Primaria	Rural	OFICIAL	Masculino	856834	107159
	Rural	NO_OFICIAL	Masculino	12419	350
	Rural	OFICIAL	Femenino	753864	64357
	Rural	NO_OFICIAL	Femenino	11444	224
	Urbana	OFICIAL	Masculino	495662	47169
	Urbana	NO_OFICIAL	Masculino	253803	10411
	Urbana	OFICIAL	Femenino	466497	31181
	Urbana	NO_OFICIAL	Femenino	228327	6592
Secundaria	Rural	OFICIAL	Masculino	175642	87844
	Rural	NO_OFICIAL	Masculino	5449	993
	Rural	OFICIAL	Femenino	158831	60233
	Rural	NO_OFICIAL	Femenino	4836	742
	Urbana	OFICIAL	Masculino	197734	84287
	Urbana	NO_OFICIAL	Masculino	79587	17804
	Urbana	OFICIAL	Femenino	194965	67880
	Urbana	NO_OFICIAL	Femenino	72101	12933

Media	Rural	OFICIAL	Masculino	58832	76054
	Rural	NO_OFICIAL	Masculino	2794	3025
	Rural	OFICIAL	Femenino	55358	66518
	Rural	NO_OFICIAL	Femenino	2486	2658
	Urbana	OFICIAL	Masculino	96562	116959
	Urbana	NO_OFICIAL	Masculino	36346	39621
	Urbana	OFICIAL	Femenino	97012	112317
	Urbana	NO_OFICIAL	Femenino	34202	36690
Ciclo 1 Adultos Ciclo 2 Adultos Ciclo 3 Adultos Ciclo 4 Adultos Ciclo 5 Adultos Ciclo 6 Adultos	Rural	OFICIAL	Masculino	116365	0
	Rural	NO_OFICIAL	Masculino	7489	0
	Rural	OFICIAL	Femenino	145716	0
	Rural	NO_OFICIAL	Femenino	6936	0
	Urbana	OFICIAL	Masculino	226466	0
	Urbana	NO_OFICIAL	Masculino	122463	0
	Urbana	OFICIAL	Femenino	279210	0
	Urbana	NO_OFICIAL	Femenino	113582	0

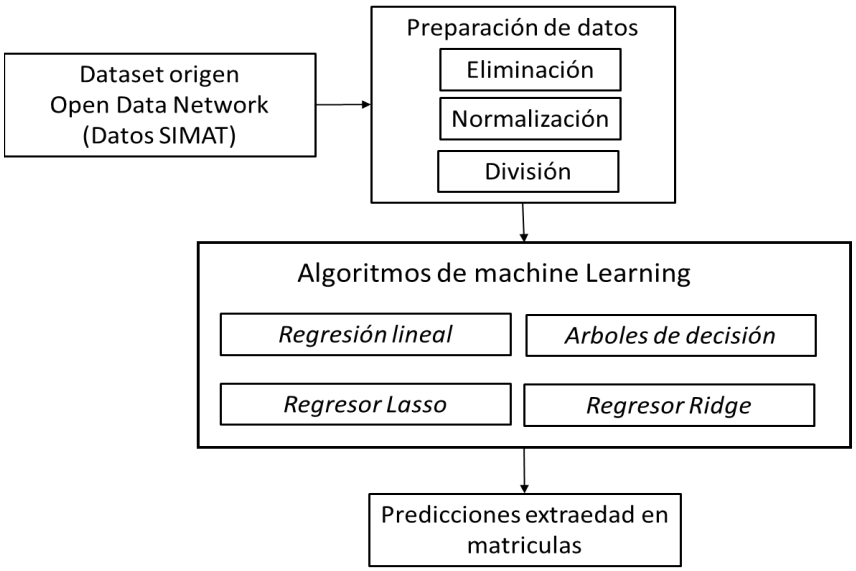
Fuente: Elaboración propia

Ante las alarmantes cifras de extraedad, toma importancia y se hace necesario un modelo que utilice información de las matrículas de todas las instituciones de educación en Colombia, con el fin de generar predicciones de las tendencias en intención matricula y acceso a la educación en zonas rurales y urbanas, dependiendo del sector oficial o no oficial y esto complementado con el género. Este modelo puede servir de base para la toma de decisiones por parte del MEN en estrategias que pueden ser planteadas como acciones administrativas que son urgentes en los procesos de dirección en los Establecimientos Educativos [18], en unión con múltiples roles (locales, territoriales y nacionales) para establecer una propuesta educativa enfocada a la población en extraedad. Además, orientaciones pedagógicas relacionadas con herramientas pedagógicas y recursos utilizados en los procesos de enseñanza y aprendizaje.

Propuesta

El modelo que se propone está basado en inteligencia artificial utilizando técnicas de machine learning. El modelo implementa cuatro técnicas a saber: Regresión lineal, Arboles de decisión (Randon Forest), regresor Lasso, regresor Rigde, específicamente orientadas a problemas de regresión. (Ver Figura 1).

Figura 1. Modelo propuesto



Elaboración propia

1. Los datos

El dataset fue obtenido desde Open Data Network, plataforma que aloja los datos abiertos del gobierno colombiano. Los datos contienen las estadísticas de la matrícula de Educación Preescolar Básica y Media de Colombia del corte 2018 a los 2020. El dataset fue validado en relación de los registros de la matrícula de educación preescolar, básica y media (EPBM), reportado en el Sistema de Matriculas Estudiantil (SIMAT) por las entidades territoriales. El dataset contiene 6916617 y 34 columnas.

2. Preparación de datos

Este se llevó a cabo siguiendo la metodología CRIPS-DM [19] e inició con un análisis exploratorio de datos con el fin de obtener un entendimiento de los datos; en este proceso se removieron columnas o variables que no aportaban a la solución, se eliminaron datos duplicados, los valores faltantes fueron completados con el promedio entre el valor anterior y el siguiente en cada columna; después se eliminaron aquellos registros que aún contenían valores nulos; por último, en el caso de la regresión, los datos originales se normalizaron con el método Min-Max, transformando los valores en un rango entre cero y uno. Se analizaron las distribuciones de las variables, los patrones que presentaban, y se identificó como se relacionaban las variables entre sí. Seguidamente, se eliminaron algunas variables tales como: *sector_conpes, cod_sector_conpes, cod_grupo_etnico, cod_especialidad, especialidad, cod_metodologia, metodología.*

Los modelos de regresión se desarrollaron en el lenguaje de programación Python utilizando las librerías *scikitlearn, pandas, numpy y matplotlib*. En el tratamiento de los datos para análisis, se empleó él método RobustScaler para normalizarlos y evitar que los resultados de los algoritmos sean afectados por valores atípicos, asimismo, los datos se dividieron en entrenamiento y pruebas, estableciendo el 30% para pruebas y 70% para entrenamiento. En el modelo se implementaron 4 técnicas de aprendizaje automático como se mencionó anteriormente: Regresión lineal múltiple, Arboles de decisión, Regresor Lasso y Regresor Rigde.

3. Algoritmos de machine Learning

El ajuste de hiperparámetros se realizó con la biblioteca *scikitlearn*, utilizando la búsqueda aleatoria (*RandomizedSearchCV*), dado que, es posible obtener resultados tan precisos como los conseguidos con la búsqueda en cuadrícula (*GridSearchCV*), aunque, con una importante reducción de tiempo, debido al muestreo de los hiperparámetros en la distribución definida, también se utilizó la validación cruzada (*RepeatedKFold*) para mejorar el rendimiento estimado de cada modelo y evitar el sobre entrenamiento. Los datos se dividieron aleatoriamente en subconjuntos y se optimizaron con la función de pérdida: error cuadrático medio.

- *Regresión lineal*, se analizó con las opciones de True/False, para utilizar o quitar la constante β_0 del regresor, se normalizo los datos para que todas las variables estén a la misma escala.

Arboles de decisión, se experimentó con distintos valores en el número de árboles, el número de características a considerar en cada división, el número máximo de niveles en el árbol, el número mínimo de muestras requeridas para dividir un nodo, el número pequeño de muestras necesarias en cada nodo hoja y el método de selección de muestras para entrenar cada árbol [20].

Regresor Lasso, Se analizó la suma de los valores absolutos de los pesos para la penalización para el parámetro *n_samples*, que es el número de observaciones para analizar el rendimiento del regresor [21].

Regresor Ridge, se evaluó el Alpha con valor de 1 equivalente a un mínimo cuadrado ordinario, además la tolerancia para la optimización con valores pequeños, la semilla del generador de números pseudoaleatorios fue tomada como 'aleatoria', para generar un coeficiente aleatorio en cada iteración. Para aumentar la varianza de las estimaciones se usó para el hiperparámetros Alpha un valor positivo de 1, por otra parte, se definió el número máximo de 15000 iteraciones para el solucionador de gradiente conjugado. En relación al parámetro *solver* se utilizó el solucionador automáticamente en función del tipo de datos [22].

Resultados

En el proceso de evaluación del modelo propuesto, fueron analizados los resultados de los años 2018 a 2022 por separado utilizando las métricas [23], coeficiente de determinación (R2), raíz del error cuadrático medio (RMSE) y error porcentual absoluto medio (MAPE). La Tabla 2, expone las ecuaciones, descripción y criterio de desempeño de cada métrica de evaluación.

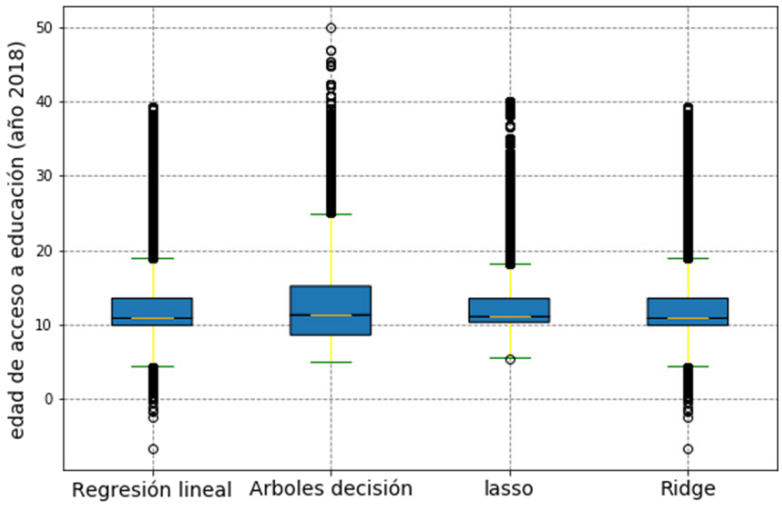
Tabla 2. Métricas de evaluación

Métrica	Ecuación	Descripción	Criterio de desempeño
R2	$1-\frac{\Sigma(y_i-x_i)^2}{\Sigma(x_i-\underline{x_i})^2}$	Define que tan cerda están los datos reales a la regresión lineal	Se encuentra en 0 y 1, a medida que tiende a 1 es ideal la utilidad del modelo.
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)^2}$	Establece la discrepancia entre los valores reales y los calculados por el modelo.	Valor mayor que 0, en cuanto más cercanía positiva a 0, ideal es el resultado de la estimación.
MAPE	$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{x_i-y_i}{x_i}\right\|\times 100$	Expresa lo bueno que puede ser la predicción del modelo, calcula el volumen de los desaciertos en las predicciones como un porcentaje.	Entre más cercanía a 0, ideal será el rendimiento del modelo.

Fuente: Elaboración propia

Para representar los resultados de las predicciones se utilizaron las gráficas de cajas y bigotes, ya que estas permiten representar y contrastar la organización y la disposición central de valores numéricos mediante sus cuartiles, con el fin de dividir los valores de la edad de matrícula en cuatro (Transición, Primaria, Secundaria y Media). Según los resultados obtenidos, se puede evidenciar que los árboles de decisión obtienen los mejores resultados, ya que la distribución de las predicciones se encuentra clara mente definidas dentro de los cuartiles, a diferencia del resto de regresores que contienen valores a típicos por encima y por debajo del rango de los cuartiles. En ese sentido, para las matrículas del año 2018, la Figura 2 muestra que los datos de extraedad para matriculas de transición inician en promedio alrededor de los 6 años, para los árboles de decisión, se encuentran en el primer cuartil, a diferencia de la regresión lineal y Ridge que contienen un número considerable de valores por debajo de este cuartil. De la misma forma para primaria, secundaria, los árboles de decisión ubican las predicciones en los cuartiles 2 y 3, con unos rangos de extra edad 8 y 16. Para el caso de la media, al igual que lo anterior, los árboles de decisión los datos se en el cuartil 4, con un rango de valores de extraedad de 17 a 25 años para la educación media, a diferencia del resto de predictores como Lasso y Ridge que colocan los rangos de extraedad en los mismos cuartiles, pero con edades entre 11 y 14 años. En relación a los datos que están por encima del cuartil 4, o en el máximo del diagrama, estos datos representan los estudiantes matriculados a Ciclo 1 Adultos, Ciclo 2 Adultos, Ciclo 3 Adultos, Ciclo 4 Adultos, Ciclo 5 Adultos, Ciclo 6 Adultos, donde la extraedad está de 25 años en adelante.

Figura 2. Predicciones de extraedad 2018



Fuente: Elaboración propia

En relación a los resultados de evaluación a la exactitud predictiva del modelo con las métricas de evaluación mencionadas. **La tabla 3** presenta los resultados, en estos se tiene que para MAPE, los árboles de decisión obtienen el mayor nivel, debido a que su valor se acerca más a cero, obteniendo un promedio de ventaja de 60% sobre el resto de los predictores. En el mismo sentido para RMSE, los árboles de decisión obtienen los mejores resultados con un promedio de ventaja de 48% en la exactitud de las predicciones. Para R2, al igual que las anteriores arboles de decisión obtiene los mejores resultados, obteniendo un 0.74233643, valor con mayor proximidad a 1. Como se puede observar en la gráfica anterior, los árboles de decisión obtienen los mejores resultados, ya que los valores de las predicciones tienen alto nivel de cercanía con la realidad que se presenta en los datos de matrícula analizados.

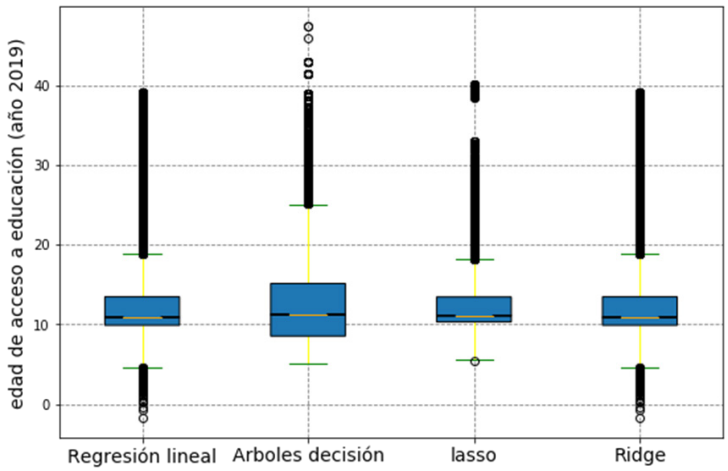
Tabla 3. Resultados métricos de evaluación año 2018

Algoritmo	MAPE	RMSE	R2
Regresión lineal	3.26392183	25.09861045	0.48888474
Árboles de decisión	1.97908828	12.65271861	0.74233643
Lasso	3.27095222	26.14351031	0.4676061
Ridge	3.2639238	25.09861022	0.48888474

Fuente: Elaboración propia

Con respecto a los datos de la predicción del 2019 estos guardan similitud con las del 2018, la Figura 3, muestra que se tiene una la ligera variación en los estudiantes de secundaria, ésta redujo pasando de 111231 en 2018 a 110863 en 2019, en relación a la educación media, al contrario, aumento de 22868 en 2018 a 23560 en 2019. Igualmente, se presenta variación en los matriculados a Ciclo 1 Adultos, Ciclo 2 Adultos, Ciclo 3 Adultos, Ciclo 4 Adultos, Ciclo 5 Adultos, Ciclo 6 Adultos, donde la extraedad está de 25 años en adelante y hasta 47 años, en lo cual los árboles de decisión son los que obtienen mejor resultado según los datos originales.

Figura 3. Predicciones de extraedad 2019



Fuente: Elaboración propia

Los resultados de la exactitud predictiva del modelo para el año 2019, se presentan en la **tabla 3**, al igual que en el 2018 los árboles de decisión obtienen los mejores valore de MAPE, notándose una pequeña mejoría a diferencia con el valor del 2018, pasando de 1.979 a 1.964, como se puede observar tiende más 0, esto se debe a que los datos de extra edad bajaron para estos dos años. De la misma forma para RMSE, los árboles de decisión obtienen los mejores resultados pasando de 12.65 en 2018 a 12.55 en 2019 igualmente este valor se acerca más a 0, Para R2, al igual que las anteriores arboles de decisión obtiene los mejores resultados, pasando de 0.735 en 2018 a 0.742 en 2019, valor con mayor proximidad a 1. Como se puede observar el número de estudiantes en extra edad tiende a bajar en la educación secundaria y a subir en la media.

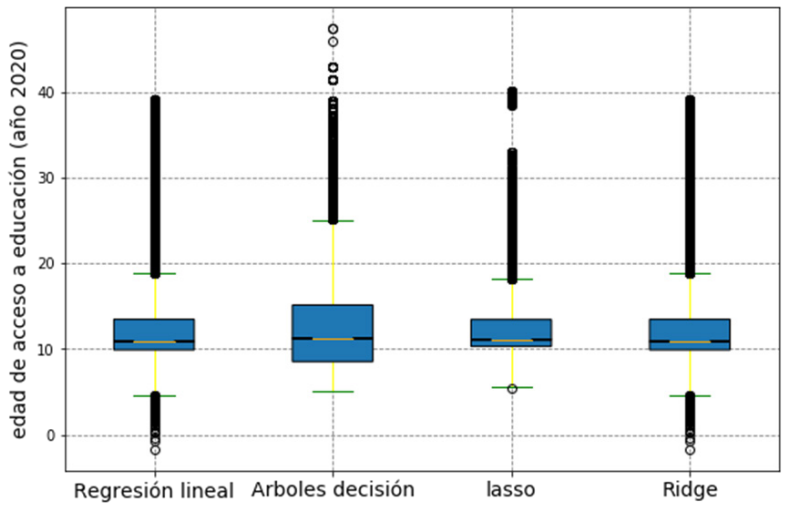
Tabla 4. Resultados métricos de evaluación año 2019

Algoritmo	MAPE	RMSE	R2
Regresión lineal	3.21442946	24.89777466	0.47501603
Arboles de decisión	1.96416975	12.55403455	0.73529093
Lasso	3.22182466	25.82901028	0.45538039
Ridge	3.2144292	24.89777464	0.4750160

Fuente: Elaboración propia

En la figura 4, se presentan los resultados de las predicciones para el año 2020, en estos al igual que los anteriores los árboles de decisión tienen la distribución con mayor relación a los datos reales, ya que los no tiene valores atípicos por debajo del cuartil 1, además en los datos del nivel máximo, los cuales están por encima del cuartil 4, pertenecen a los estudiantes de Ciclo 1 Adultos, Ciclo 2 Adultos, Ciclo 3 Adultos, Ciclo 4 Adultos, Ciclo 5 Adultos, Ciclo 6 Adultos. En relación a los estudiantes de Transición en extraedad se tiene que bajo de 9594 en 2019 a 7767 en 2020, Según [1], la disminución de la extraedad en los estudiantes de transición se debe a la implementación de metodologías, actividades, estrategias didácticas, la intensidad de la jornada escolar, el horario, los tiempos previstos para lograr las metas de aprendizaje, los recursos pedagógicos y el proceso de evaluación formativa, que se ha ido implementado a nivel nacional. Por otra en los estudiantes de primaria igualmente disminuyo el numero pasando de 48552 en 2019 a 43974 en 2020, estos datos corroboran que las estrategias planeadas por el MEN están funcionando de alguna forma. En relación a secundaria la disminución fue un poco leve paso de 110863 en 2019 a 110622 en 2020, aunque el MEN ha planteado estrategias de nivelación de estudiantes de la zona rural y urbana que permitan garantizar la permanencia y reingreso de aquellos estudiantes que por encontrarse en condición de extraedad, en algunos casos estos siguen abandonando el sistema educativo. De forma contraria ocurre con los estudiantes de media donde el número de estudiantes en extraedad aumento paso de 23560 en 2019 a 23822 en 2020, los posibles orígenes apuntan a los problemas sociales y económicos, entre los que destacan, costo de oportunidad, de transporte y de alimentación, por estos motivos los estudiantes de esta etapa de educación prefieren trabajar.

Figura 4. Predicciones de extraedad 2020



Fuente: Elaboración propia

Los resultados de la exactitud predictiva del modelo para el año 2020, se presentan en la tabla 5, al igual que en 2018 y 2019, los árboles de decisión obtienen los mejores valore de MAPE, el modelo mejoro es decir aumento su precisión, alcanzando un valor de 1.746, es decir cada vez este valor tiende más a 0, métrica que corrobora lo ocurrido, es decir la disminución de estudiantes en extraedad entre los años 2018 y 2020. En el mismo sentido, para RMSE, los árboles de decisión obtienen los mejores resultados logrando un valor de 9.930 menor que los valores alcanzados en esta métrica para los años 2018 y 1019, lo cual demuestra que el modelo ha mejorado su exactitud en relación a los dos años pasados. Para R2, al igual que las anteriores arboles de decisión logro un valor de 0.738 más aproximado a 1. Estos resultados confirman que las estrategias planteadas en [1], de alguna forma están funcionando.

Tabla 5. Resultados métricos de evaluación año 2020

Algoritmo	MAPE	RMSE	R2
Regresión lineal	2.9983963	21.77780107	0.42702153
Árboles de decisión	1.7460446	9.93083546	0.73871766
Lasso	2.99552609	22.80870504	0.39989823
Ridge	2.99839587	21.77780118	0.42702153

Fuente: Elaboración propia

Conclusiones

En este artículo se utilizaron algoritmos de machine Learning una rama de la inteligencia artificial para predecir las tendencias en el problema de matrícula de estudiantes en extraedad en la educación básica en Colombia. El dataset descargado desde Open Data Network, dio la posibilidad de trabajar con directamente las estadísticas reales de la matrícula de Educación Preescolar Básica y Media en Colombia. En este sentido según los resultados obtenidos por el modelo, este pude servir de base en el ministerio de educación para la implementación de políticas de orientación administrativas para la atención educativa de la población en extraedad, joven, adulta y adulta mayor con discapacidad intelectual y psicosocial, en el marco de la inclusión y la equidad en la educación tanto rural como urbana.

El modelo permitió identificar que unos de los factores de mayor relevancia para que un estudiante se abandone sus estudios, es el económico, ya que en muchos casos los alumnos tengan que trabajar aportar con la superación de las dificultades financieras en el hogar. Además, problemas sociales y económicos, debido a los costos, en algunos casos, vale la pena más que el niño o joven trabaje, que lo que cuesta mandarlo a la escuela, por otra parte, obstáculos cotidianos tales como: no pueden pagar transporte, comida, útiles, en consecuencia, para algunos estudiantes, la educación dejó de ser prioridad.

Los resultados de los estudiantes en cada uno de los niveles de educación básica permiten identificar qué es necesario una educación inclusiva donde todos los niños, niñas, adolescentes, jóvenes y adultos, según sus necesidades, intereses, posibilidades y expectativas, independientemente de su género, discapacidad, capacidad o talento excepcional, pertenencia étnica, posición política, ideología, asisten y participan de una educación en la que comparten con pares de su misma edad y reciben los apoyos que requieren para que su educación sea exitosa [24].

Los resultados, dejaron en evidencia que de alguna manera las políticas implementadas en relación a estrategias pedagógicas, estrategias didácticas, la intensidad de la jornada escolar, han contribuido a la disminución del número de estudiantes en extraedad en los niveles de transición y primaria, ya que entre el 2018 y 2020 el número de estudiantes en extraedad bajo en los sectores rural y urbano.

Como trabajos futuros, se espera poder contar con mayor cantidad de datos, los cuales pueden cubrir años 2021 y 2022. Además, se espera poder aplicar una técnica de algoritmos de machine Learning denominada ensamble de modelos, la cual permite aumentar la precisión y mejorar los resultados del modelo.

Referencias Bibliográficas

1. S. Gupta and M. K. Jawanda, "The impacts of COVID-19 on children," *Acta Paediatr. Int. J. Paediatr.*, vol. 109, no. 11, pp. 2181–2183, 2020. DOI: 10.1111/apa.15484.
2. J. M. Mirasol, J. V. Belderol Necosia, B. B. Bicar, and H. P. Garcia, "Statutory policy analysis on access to Philippine quality basic education," *Int. J. Educ. Res. Open*, vol. 2, no. November, p. 100093, 2021. DOI: 10.1016/j.ijedro.2021.100093.
3. B. Hunter *et al.*, "Strengthening global midwifery education to improve quality maternity care: Co-designing the World Health Organization Midwifery Assessment Tool for Education (MATE)," *Nurse Educ. Pract.*, vol. 63, no. April, 2022. DOI: 10.1016/j.nepr.2022.103376.
4. ONU, "Educación de calidad: por qué es importante," *Un.Org*, pp. 1–6, 2017, [Online]. Available: http://www.un.org/%0Ahttp://www.un.org/sustainabledevelopment/es/wp-content/uploads/sites/3/2016/10/4_Spanish_Why_it_Matters.pdf
5. UNESCO, "Inform: Education in Colombia," 2016, [Online]. Available: https://www.oecd.org/education/school/Education-in-Colombia-Highlights.pdf
6. S. Schelfhout *et al.*, "How accurately do program-specific basic skills predict study success in open access higher education?," *Int. J. Educ. Res.*, vol. 111, no. November 2021, p. 101907, 2022, doi: 10.1016/j.ijer.2021.101907.
7. Ministerio de educaciòn, "GUIDE OF THE COLOMBIAN EDUCATIONAL SYSTEM AND ASPECTS TO BE CONSIDERED WHEN UNDERTAKING HIGHER EDUCATION STUDIES IN COLOMBIA," no. 3. pp. 1–27, 2018.
8. M. Delgado Barrera, "La Educación Básica y Media en Colombia Retos en Equidad y Calidad," *Fedesarrollo Cent. Investig. Económica y Soc.*, pp. 1–40, 2014, [Online]. Available: https://www.repository.fedesarrollo.org.co/handle/11445/190%0Ahttp://hdl.handle.net/11445/190
9. D. MORÓN and L. Pachano, "La extraedad como factor de segregación y exclusión escolar*," *Rev. Pedagog.*, vol. 27, pp. 33–69, 2006.
10. S. Rodriguez-Raga and N. Martinez-Camelo, "Game, guide or website for financial education improvement: Evidence from an experiment in Colombian schools," *J. Behav. Exp. Financ.*, vol. 33, p. 100606, 2022. DOI: 10.1016/j.jbef.2021.100606.
11. M. Wahl and D. Majchrzak, "The impact of a sensory education on gustatory and olfactory perception in Austrian school children aged 11 to 14 – A consideration of long-term effects," *Food Qual. Prefer.*, vol. 98, no. July 2021, p. 104527, 2022. DOI: 10.1016/j.foodqual.2022.104527.
12. C. Ordoñez, E. Ruano, C. Cobos, H. Ordoñez, and A. Ordoñez, "Comparative Analysis of MOGBHS with Other State-of-the-Art Algorithms for Multi-objective Optimization Problems BT - Advances in Soft Computing," 2018, pp. 154–170.
13. y A. A. M. M. E. J. De la Hoz Domínguez, T. J. Fontalvo Herrera, "Aprendizaje automático y PYMES: Oportunidades para el mejoramiento del proceso de toma de decisiones," *Investig. e Innovación en Ing.*, vol. 8, no. 1, pp. 21–36, 2020.
14. E. D. Smolyansky, H. Hakeem, Z. Ge, Z. Chen, and P. Kwan, "Machine learning models for decision support in epilepsy management: A critical review," *Epilepsy Behav.*, vol. 123, 2021. DOI: 10.1016/j.yebeh.2021.108273.

15. H. Ordóñez, C. Cobos, and V. Bucheli, “Machine learning model for predicting theft trends in Colombia,” *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2020, no. E29, pp. 494–506, 2020.

16. O. Embarak, “A New Paradigm through Machine Learning: A Learning Maximization Approach for Sustainable Education,” *Procedia Comput. Sci.*, vol. 191, pp. 445–450, 2021. DOI: 10.1016/j.procs.2021.07.055.

17. A. Bertoletti, J. Berbegal-Mirabent, and T. Agasisti, “Higher education systems and regional economic development in Europe: A combined approach using econometric and machine learning methods,” *Socioecon. Plann. Sci.*, no. November 2020, p. 101231, 2022. DOI: 10.1016/j.seps.2022.101231.

18. A. Marcela and O. Vera, *Orientaciones administrativas y pedagógicas para la atención educativa de la población en extraedad, joven, adulta y adulta mayor con discapacidad intelectual y psicosocial*. 2018. [Online]. Available: https://www.colombiaaprende.edu.co/sites/default/files/files_public/2021-07/Documento de Orientaciones para extraedad con ISBN %281%29.pdf

19. J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, “Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test,” *Procedia Comput. Sci.*, vol. 198, no. 2020, pp. 512–517, 2021. DOI: 10.1016/j.procs.2021.12.278.

20. M. Moshkov, “Decision trees for regular factorial languages,” *Array*, vol. 15, no. January, p. 100203, 2022. DOI: 10.1016/j.array.2022.100203.

21. F. Centofanti, M. Fontana, A. Lepore, and S. Vantini, “Smooth LASSO estimator for the Function-on-Function linear regression model,” *Comput. Stat. Data Anal.*, vol. 176, p. 107556, 2022. DOI: 10.1016/j.csda.2022.107556.

22. W. Jamil and A. Bouchachia, “Iterative ridge regression using the aggregating algorithm,” *Pattern Recognit. Lett.*, vol. 158, pp. 34–41, 2022, DOI: 10.1016/j.patrec.2022.04.021.

23. C. Pardo, E. Suescún, H. Jojoa, R. Zambrano, y W. Ortega, "Modelo de referencia para la adopción e implementación de Scrum en la industria de software", Investigación e Innovación en Ingenierías, vol. 8, n.º 3, pp. 14-28, 2020. https://doi.org/10.17081/invinno.8.3.4700

24. J. Dessain, “Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric,” *Expert Syst. Appl.*, vol. 199, no. March, p. 116970, 2022. DOI: 10.1016/j.eswa.2022.116970.