



## UNIVERSIDAD TECNOLÓGICA ISRAEL

### ESCUELA DE POSGRADOS “ESPOG”

### MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

*Resolución: RPC-SO-32-No.536-2023*

#### PROYECTO DE TITULACIÓN EN OPCIÓN AL GRADO DE MAGISTER

|  |
|--|
| <b>Título del proyecto:</b>  |
| Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad<br>Inconclusa en Estudiantes del Ministerio de Educación |
| <b>Línea de Investigación:</b>   |
| Ciencias de la ingeniería aplicadas a la producción, sociedad y desarrollo sustentable   |
| <b>Campo amplio de conocimiento:</b>   |
| Tecnologías de la Información y la Comunicación (TIC)  |
| <b>Autor:</b>  |
| Ale x Paul Yugsi Cayambe   |
| <b>Tutores:</b>  |
| Renato Mauricio Toasa Guachi<br>Mario Rubén Pérez Cargua   |

Quito – Ecuador

2025

## APROBACIÓN DEL TUTOR



Yo, Renato Mauricio Toasa Guachi con C.I: 1804724167 en mi calidad de Tutor del proyecto de investigación titulado: Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación.

Elaborado por: Alex Paúl Yugsi Cayambe, de C.I: 1725056475, estudiante de la Maestría: Big Data y Ciencia de Datos, de la **UNIVERSIDAD TECNOLÓGICA ISRAEL (UISRAEL)**, como parte de los requisitos sustanciales con fines de obtener el Título de Magister, me permito declarar que luego de haber orientado, analizado y revisado el trabajo de titulación, lo apruebo en todas sus partes.

Quito D.M., marzo de 2025

---

**Firma**

## APROBACIÓN DEL TUTOR



Yo, Mario Rubén Pérez Cargua con C.I: 0603251984 en mi calidad de Tutor del proyecto de investigación titulado: Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación.

Elaborado por: Alex Paúl Yugsi Cayambe, de C.I: 1725056475, estudiante de la Maestría: Big Data y Ciencia de Datos, de la **UNIVERSIDAD TECNOLÓGICA ISRAEL (UISRAEL)**, como parte de los requisitos sustanciales con fines de obtener el Título de Magister, me permito declarar que luego de haber orientado, analizado y revisado el trabajo de titulación, lo apruebo en todas sus partes.

Quito D.M., marzo de 2025

---

**Firma**

## DECLARACIÓN DE AUTORIZACIÓN POR PARTE DEL ESTUDIANTE



Yo, Alex Paúl Yugsi Cayambe con C.I: 1725056475, autor del proyecto de titulación denominado: Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación. Previo a la obtención del título de Magister en Big Data y Ciencia de Datos.

1. Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar el respectivo trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.
2. Manifiesto mi voluntad de ceder a la Universidad Tecnológica Israel los derechos patrimoniales consagrados en la Ley de Propiedad Intelectual del Ecuador, artículos 4, 5 y 6, en calidad de autor del trabajo de titulación, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital como parte del acervo bibliográfico de la Universidad Tecnológica Israel.
3. Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de prosperidad intelectual vigentes.

Quito D.M., marzo de 2025

---

**Firma**

## Tabla De Contenido

|  |    |
|--|----|
| APROBACIÓN DEL TUTOR .....                                 | 2  |
| APROBACIÓN DEL TUTOR .....                                 | 3  |
| DECLARACIÓN DE AUTORIZACIÓN POR PARTE DEL ESTUDIANTE ..... | 4  |
| INFORMACIÓN GENERAL .....                                  | 1  |
| Contextualización del tema.....                            | 1  |
| Problema de investigación.....                             | 3  |
| Objetivo general.....                                      | 4  |
| Objetivos específicos.....                                 | 4  |
| Vinculación con la sociedad y beneficiarios directos:..... | 4  |
| CAPÍTULO I: DESCRIPCIÓN DEL PROYECTO .....                 | 5  |
| 1.1.    Contextualización general del estado del arte..... | 5  |
| 1.2.    Proceso investigativo metodológico .....           | 7  |
| 1.3.    Análisis de resultados .....                       | 8  |
| CAPÍTULO II: PROPUESTA.....                                | 16 |
| 1.1.    Fundamentos teóricos aplicados.....                | 16 |
| 1.2.    Descripción de la propuesta .....                  | 18 |
| 1.3.    Validación de la propuesta .....                   | 27 |
| 1.4.    Matriz de articulación de la propuesta .....       | 29 |
| CONCLUSIONES.....  | 31 |
| RECOMENDACIONES.....                                       | 32 |
| BIBLIOGRAFÍA.....  | 33 |
| ANEXOS .....   | 35 |
| ANEXO 1 .....  | 35 |
| ANEXO 2 .....  | 39 |
| ANEXO 3 .....  | 41 |
| ANEXO 4 .....  | 43 |

## Índice de tablas

|  |    |
|--|----|
| Tabla 1 Elección de Herramientas de Google Cloud para Machine Learning ..... | 25 |
| Tabla 2 Costo – Beneficio Google Cloud .....                                 | 26 |
| Tabla 3 Matriz de articulación.....  | 29 |

## Índice de figuras

|   |    |
|---|----|
| Figura 1 Pregunta 1 .....                                       | 8  |
| Figura 2 Pregunta 2 .....                                       | 9  |
| Figura 3 Pregunta 3 .....                                       | 9  |
| Figura 4 Pregunta 4 .....                                       | 10 |
| Figura 5 Pregunta 5 .....                                       | 10 |
| Figura 6 Pregunta 6 .....                                       | 11 |
| Figura 7 Pregunta 7 .....                                       | 11 |
| Figura 8 Pregunta 8 .....                                       | 12 |
| Figura 9 Pregunta 9 .....                                       | 12 |
| Figura 10 Pregunta 10 .....                                     | 13 |
| Figura 11 Pregunta 11 .....                                     | 13 |
| Figura 12 Pregunta 12 .....                                     | 14 |
| Figura 13 Pregunta 13 .....                                     | 14 |
| Figura 14 Pregunta 14 .....                                     | 15 |
| Figura 15 Pregunta 15 .....                                     | 15 |
| Figura 16 Estructura general de la propuesta.....               | 19 |
| Figura 17 Estructura de datos del ministerio de educación. .... | 20 |
| Figura 18 Datos históricos de estudiantes. ....                 | 20 |
| Figura 19 Proyección abandono de estudiantes .....              | 23 |

## INFORMACIÓN GENERAL

### Contextualización del tema

La educación constituye una pieza clave para el desarrollo social y económico de una nación, en Ecuador, a pesar de los esfuerzos del Ministerio de Educación (MINEDUC) por mejorar el acceso y la calidad educativa, también persisten limitaciones respecto a la permanencia y finalización de los estudios; un porcentaje elevado de estudiantes abandona la educación sin culminar y se conoce como escolaridad inconclusa. Esta condición tiene efectos negativos sobre el acceso a los trabajos y refuerza la pobreza y desigualdad social (Mejía, 2023).

A modo de solución, el uso de tecnologías avanzadas como son el Machine Learning o Data Science puede ser una alternativa innovadora para hacer uso de grandes volúmenes de datos e identificar patrones de riesgo de deserción escolar; los modelos predictivos que se pueden implementar permitirían identificar tempranamente a los estudiantes vulnerables y desarrollar acciones de intervención más adecuadas que mejoren la retención escolar (Benalcázar, 2024)

La presente investigación propone explorar la aplicación de un modelo con Machine Learning en el MINEDUC para analizar la escolaridad inconclusa; tal investigación tiene el propósito de constituir un conocimiento útil para una toma de decisiones en el ámbito educativo y, finalmente, proponer acciones que favorezcan la disminución de este problema, como entregar mejores oportunidades de aprendizaje y desarrollo a los estudiantes del Ecuador.

### **1.2. Análisis de la Escolaridad Inconclusa: Situación Actual en el MINEDUC**

Un importante reto presenta el país, en cuanto a la reducción de la deserción, el MINEDUC estima que un 20% de los estudiantes abandona la educación formal antes de finalizar la educación media, un porcentaje que en las zonas rurales se incrementa a más del 30% en las zonas urbano-marginales (Mejía, 2023), se trata de un fenómeno que pone en evidencia las desigualdades todavía existentes en el país en el acceso y calidad educativa.

La implementación de programas de inclusión y becas por parte del gobierno no ha sido tampoco suficiente para hacer frente a este problema; en este sentido, es posible concluir que la aplicación de Data Science y Machine Learning juega un papel clave para comprender y ayudar a mitigar la deserción escolar; estos elementos permiten realizar analíticas más precisas y con mayor profundidad, determinar las características que definen el abandono escolar y que sirven de base para las políticas educativas basadas en datos (Benalcázar, 2024).

Así, el uso de modelos predictivos por parte del MINEDUC permitiría anticipar qué estudiantes tienen mayor riesgo de abandonar la educación escolar y poder implementar, por tanto, acciones que trataran de evitar el abandono, esta práctica podría también ayudar a optimizar la utilización de los recursos y, sobre todo, contribuir a incrementar la permanencia escolar en el país.

### **1.3 Análisis de los Factores**

La escolaridad inconclusa en Ecuador es un fenómeno con múltiples determinantes que abarca las variables socioeconómicas, culturales, familiares y del sistema educativo; entre los factores más críticos se pueden resaltar:

**Factores socioeconómicos:** La pobreza y la presión por tener que trabajar temprano hacen que los estudiantes no continúen su escolaridad (Jaramillo, 2022).

**Factores culturales:** En determinadas comunidades se considera que la educación no es fundamental y se perpetúan los roles de género, lo que impide que las niñas y adolescentes accedan a la escolaridad (Monteros, 2024).

**Factores familiares:** La ausencia de núcleos familiares, problemas de violencia, la migración, entre otros, exacerban los niveles de deserción escolar (Tumbaco, 2022).

**Factores del sistema educativo:** La infraestructura educativa inadecuada, el acceso escaso a la tecnología y los programas de apoyo deficientes, como las tutorías, tienen un efecto en la deserción escolar (Jaramillo, 2022), un modelo de Machine Learning podría incorporar estas variables para realizar una predicción más adecuada de qué estudiantes se encuentran en mayor riesgo de abandono de la escolaridad y poder implementar acciones para su atención.

### **1.4 Escolaridad Inconclusa y el uso del Machine Learning**

El abandono escolar tiene impactos tanto a nivel individual como colectivo. Al nivel individual, limita el acceso a empleos de calidad y perpetúa en el tiempo los ciclos de la pobreza, aumentando la vulnerabilidad social de los jóvenes (Freire, 2023); a nivel social, se agravan las desigualdades económicas y se reduce la competitividad laboral del país; además, se enfrenta a un uso ineficaz de los recursos públicos, donde las inversiones en educación no producen los efectos presuponen que se generen al tener estudiantes que no son escolarizados. En este sentido, queda evidenciada la necesidad de estrategias educativas que estén soportadas en las nuevas tecnologías para tratar la escolaridad inconclusa.

La eventual implementación de Machine Learning por el MINEDUC tiene la potencialidad de facilitar la gestión de datos educativos, lo que permitiría detectar a estudiantes con riesgo de escolaridad inconclusa de manera más temprana, así como mejorar la toma de decisiones en las políticas educativas. Mediante la predicción, se puede dar paso al desarrollo de programas de intervención para las aulas, una optimización de los recursos y contribuir a una educación más inclusiva y equitativa.

### **Problema de investigación**

El núcleo de la presente investigación trata sobre la aplicación de un modelo de Machine Learning orientado a la identificación de los factores por los que se produce la escolaridad inconclusa del sistema educativo ecuatoriano, en específico, del conjunto de alumnos y alumnas consideradas por el sistema del MINEDUC, a pesar de que el Gobierno ha puesto en marcha diferentes políticas públicas en el ámbito educativo orientadas a la reducción de la reducción de la deserción escolar, aún hay un porcentaje de estudiantes que abandona sus estudios entre completar la educación básica y la educación media, lo que perjudica el desarrollo personal de los estudiantes abandonantes y, además, limita el acceso a mejores oportunidades laborales (Mejía, 2023).

El proyecto se encuentra en búsqueda de la identificación de un modelo predictivo de abandono escolar que el MINEDUC puede utilizar de forma muy efectiva y precisa para la identificación de aquellos estudiantes que están en alto riesgo de abandono de sus estudios. La aplicación de Machine Learning pueden resultar útiles para el análisis de volúmenes de datos educativos, la identificación de patrones de abandono de la educación y, a su vez, para el diseño de las estrategias de intervención a partir de la evidencia (Benalcázar, 2024).

Como consecuencia de esta problemática surgen preguntas que pueden ser consideradas fundamentales en el ámbito del sistema educativo ecuatoriano: ¿Cuáles son aquellos factores que emergen como los principales en la deserción escolar? ¿Puede preverse con antelación cuáles de esos estudiantes están en mayor riesgo de abandonar sus estudios? y ¿puede un modelo de Machine Learning colaborar en favor de la anticipación a la deserción y a la formación de estrategias más efectivas?

## **Objetivo general**

Usar un modelo de Machine Learning para el análisis de la escolaridad inconclusa en estudiantes del Ministerio de Educación

## **Objetivos específicos**

- Contextualizar los fundamentos teóricos sobre la escolaridad inconclusa, el uso de Machine Learning en la educación y su aplicación en la detección de factores de riesgo en el abandono escolar.
- Diagnosticar los factores determinantes en la deserción escolar en Ecuador, mediante el análisis de datos proporcionados por el MINEDUC.
- Aplicar un modelo de Machine Learning que permita evaluar el riesgo de deserción escolar en estudiantes del sistema educativo ecuatoriano.
- Evaluar la propuesta mediante el criterio de especialistas.

## **Vinculación con la sociedad y beneficiarios directos:**

Este proyecto tiene un impacto en la sociedad, ya que busca reducir la escolaridad inconclusa en Ecuador mediante un modelo de Machine Learning que aplica el Ministerio de Educación, la vinculación con la colectividad se llevará a cabo mediante un análisis educacional de los datos y la aplicación de estrategias que permitan la detección anticipada de los y las jóvenes en riesgo de deserción escolar.

Como contribución a la sociedad, el modelo facilitará el incremento de la escolaridad, lo que, a la larga, podrá contribuir a la reducción de la pobreza, el desempleo y desigualdades sociales. Además, se ayudará a profundizar la equidad educativa al poder identificar con exactitud a la población estudiantil que se considere más vulnerable, de modo que los esfuerzos e insumos del MINEDUC puedan ser utilizados de una manera más eficaz.

Los beneficiarios directos que se verán beneficiados por el proyecto son los y las jóvenes en riesgo de desertar de la escuela, el Ministerio de Educación, los docentes, directivos educativos que podrán aprovechar los resultados del modelo a fin de retener a los y las jóvenes e incrementar la calidad del sistema educativo ecuatoriano.

## CAPÍTULO I: DESCRIPCIÓN DEL PROYECTO

### 1.1. Contextualización general del estado del arte

#### 1.1.1. Escolaridad inconclusa y abandono escolar.

La escolaridad inconclusa, representa una de las principales barreras al desarrollo social y económico de los países. En palabras de Mejía (2023), el abandono escolar se produce cuando los estudiantes no llegan a concluir el tramo de educación básica o media antes de abandonar definitivamente la escuela, lo que provoca una repercusión negativa en su futuro laboral, su integración a la sociedad o al bienestar general del joven pero, además, en un problema que da cuenta de la desigualdad; dado que, aquellos jóvenes que abandonan sus estudios lo harán con más dificultades en llegar a contar con un empleo formal y, posteriormente, en mejorar su calidad de vida. Se contabilizan múltiples factores que inciden con el abandono escolar; Jaramillo (2022) hace hincapié en lo socioeconómico (pobreza, el trabajo desde muy pequeño, etc.) como las principales causas, y Monteros (2024) menciona que en comunidades rurales, culturales y de género (la educación no se consideraba como prioridad en su grupo, el hecho de que, en muchas ocasiones, la educación, no es una prioridad en su grupo, pero sí lo es para las niñas y los adolescentes) son otros elementos a tener en cuenta.

Tampoco los programas gubernamentales, se han mostrado positivos en el total de su implementación, dadas las carencias detectadas en los programas de becas o ayudas económicas, los cuales sí han conseguido progresivamente una disminución del abandono escolar y la escolaridad inconclusa, sin embargo, lejos de la erradicación del abandono. Ello nos lleva a urgir una nueva forma de aproximación para poder hacer el análisis del abandono escolar y la escolaridad inconclusa, y es la del uso de nuevas tecnologías que permitan hacer análisis más profundos de los factores que están en juego en la escolaridad inconclusa, identificando patrones y estrategias de prevención.

#### 1.1.2. Aplicación del Machine Learning en el ámbito Educativo

Aprovechando el desarrollo tecnológico, se han implementado nuevas herramientas con el objetivo de mejorar la calidad educativa y la reducción de la deserción escolar; en este contexto Benalcázar (2024) sostiene que el Machine Learning ha sido una alternativa utilizada para analizar grandes volúmenes de datos con el fin de predecir tendencias en la conducta de los estudiantes; estas técnicas han sido utilizadas con éxito en distintos países del mundo para determinar factores de riesgo que pueden verse ligados a los indicadores de deserción escolar y para la creación de planes de intervención específicos.

Entre las aplicaciones más utilizadas se encuentran los modelos de predicción que son implementados a través de algoritmos Machine Learning como la regresión, este tipo de modelos analizan datos históricos de un estudiante en particular que pueden incluir el rendimiento escolar, los días de asistencia y el contexto socioeconómico del alumno para determinar su propensión a abandonar sus estudios. Tumbaco (2022) señala que combinar el Big Data y el Machine Learning son métodos que permiten resultados más precisos en la identificación de patrones en la deserción; aun así, uno de los retos que presentan estos modelos es la calidad de los datos).

Sin embargo, en este avance existen desafíos que continúan en la investigación. La aplicabilidad de estos modelos en distintos contextos educativos y su convergencia con políticas públicas son parte de esta dificultad; el presente trabajo pretende buscar esta brecha al utilizar un modelo de Machine Learning desde esta perspectiva hacia un modelo determinado para el sistema educativo ecuatoriano, y la aplicabilidad de este en el contexto del Ministerio de Educación.

### **1.1.3. Beneficios y Desafíos del Uso de Machine Learning en el MINEDUC**

El uso de un modelo de Machine Learning en el MINEDUC puede suponer una transformación importante de la gestión educativa y de la mejora de la identificación preventiva de los estudiantes en riesgo de deserción escolar; Freire (2023) habla de sistemas educativos que han incorporado tecnologías de la llamada predicción de aprendizaje y que han mejorado sus tasas de deserción escolar, lo que a su vez les otorga la capacidad de atender a los estudiantes más vulnerables antes de que abandonen la escuela; los beneficios de esta práctica son los siguientes:

- Identificación de los riesgos de forma temprana: el modelo identifica a los estudiantes en riesgo de deserción a partir del estudio de datos históricos y de factores contextuales en tiempo real.
- Eficiencia en la utilización de los recursos: facilita el hecho de que las políticas y los programas de la educación se enfoquen a los sectores donde las tasas de deserción son más altas.

Por otro lado, la adopción de Machine Learning en el MINEDUC no está exenta de desafíos. Según Benalcázar (2024), uno de los problemas es la resistencia al cambio en las instituciones educativas, donde persisten los mecanismos de análisis y evaluación tradicionales; tampoco existe en la actualidad una normativa o regulación que garantice la privacidad y la seguridad de los datos extraídos de los estudiantes en la práctica educativa; frente a estos desafíos, es

necesario contar con un enfoque colaborativo entre el MINEDUC, las instituciones educativas y los expertos en inteligencia artificial; el objetivo de la investigación que aquí se presenta no está sólo orientado a aplicar un modelo de predicción que funcione, sino que también se centra en la viabilidad de su exploración, es decir, en la implementación y sostenibilidad de su aplicación en el sistema educativo ecuatoriano.

## **1.2. Proceso investigativo metodológico**

### **1.2.1. Enfoque de la Investigación**

La investigación que aquí se presenta adoptará un enfoque cuantitativo puesto que persigue hacer una encuesta de quince preguntas cerradas con el objetivo de recolectar y analizar datos numéricos que faciliten el desarrollo de un modelo de predicción de escolaridad inconclusa a partir de la técnica del Machine Learning, este enfoque generará información numérica que permite medir con precisión la relación entre variables y predecir la deserción escolar existente en función de los datos históricos obtenidos a través de datos del Ministerio de Educación (MINEDUC).

Para la recopilación de datos, se aplicó un muestreo no probabilístico por conveniencia, encuestando a diecinueve personas seleccionadas por el conocimiento del tema. La encuesta utilizada se encuentra en el **Anexo 1**.

### **1.2.2. Tipo de Investigación**

Este estudio se desarrollará bajo un diseño exploratorio, descriptivo y correlacional.

En la primera fase del estudio, el carácter exploratorio permitirá analizar los factores que contribuyen a la escolaridad inconclusa en Ecuador, identificando variables relevantes sin establecer relaciones causales definitivas; posteriormente, en la fase descriptiva, se llevará a cabo un análisis detallado de los datos abiertos del Mineduc, con el propósito de caracterizar las principales variables que influyen en el abandono escolar y determinar su distribución en la población estudiantil, finalmente, el estudio adoptará un enfoque correlacional, en el cual se aplicará Machine Learning para establecer relaciones entre las variables estudiadas y predecir la probabilidad de deserción escolar, este enfoque metodológico permitirá comprender el problema en profundidad y aplicar una herramienta analítica basada en datos, que ayude a reducir la deserción escolar en Ecuador mediante estrategias de intervención temprana.

### 1.3.3. Métodos, Técnicas e Instrumentos

La investigación llevará en primer lugar, el método científico cuantitativo como método de investigación, el cual estará orientado a técnicas de recolección y análisis de datos mediante 19 encuestas.

La recolección de datos para la investigación será uno de los primeros pasos de esta investigación. En esta etapa se realizaría la obtención de aquellos datos de alumnos con periodos históricos, los cuales ya han proporcionados a la investigación MINEDUC con aquellas variables sociodemográficas, académicas y familiares. La muestra de los datos en la investigación será la recolección de los datos, mediante técnicas de depuración y normalización de datos.

### 1.3. Análisis de resultados

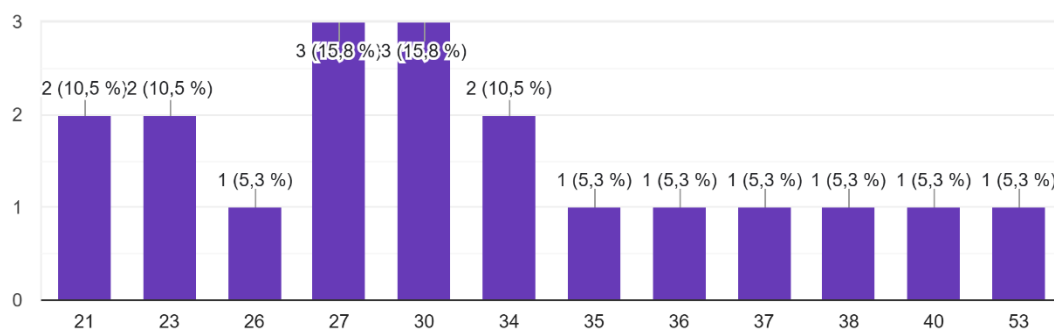
A continuación, se presentan los resultados obtenidos en la encuesta aplicada a los participantes de la investigación. Este proceso permitió recopilar información relevante sobre las características y opiniones de los encuestados, lo cual servirá de base para interpretar los factores relacionados con la escolaridad inconclusa. Cada gráfico será analizado de manera individual, utilizando un lenguaje claro y sencillo, para comprender mejor el perfil y la situación de los estudiantes.

**Figura 1**

#### Pregunta 1

¿Cuál es su edad?

19 respuestas



En este gráfico se observa que participaron 19 personas en la encuesta. Las edades son variadas, pero hay grupos que destacan. Por ejemplo, las edades de 27 y 30 años son las más comunes, ya que cada una tiene 3 personas, lo que representa 15,8% de los encuestados.

También hay varios participantes de 21, 23 y 34 años, con 2 personas en cada grupo, que representan 10,5% cada uno. En cambio, hay edades menos frecuentes como 26, 35, 36, 37,

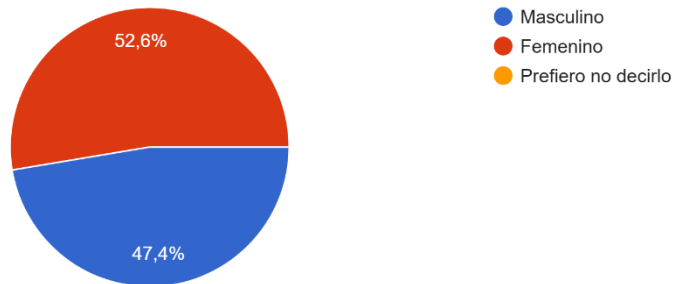
38, 40 y 53 años, donde solo hay 1 persona por cada edad, lo que equivale a 5,3% en cada caso.

**Figura 2**

*Pregunta 2*

¿Cuál es su género?

19 respuestas



Se observa que el 52,6% corresponde a personas de género femenino, mientras que el 47,4% son de género masculino.

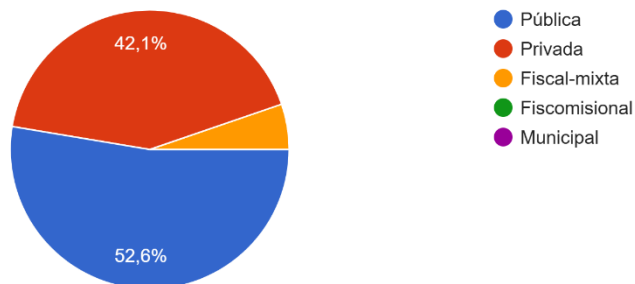
Esto indica que hay una participación equilibrada entre hombres y mujeres, aunque hay una ligera mayoría de mujeres. En este caso, ninguna persona seleccionó la opción "Prefiero no decirlo", lo cual muestra que los participantes se sintieron cómodos al responder esta pregunta.

**Figura 3**

*Pregunta 3*

¿En qué tipo de institución estudia/estudió?

19 respuestas



En este gráfico se observa que el 52,6% de los encuestados estudia o estudió en una institución pública, lo cual muestra que la mayoría accedió a la educación gratuita del Estado.

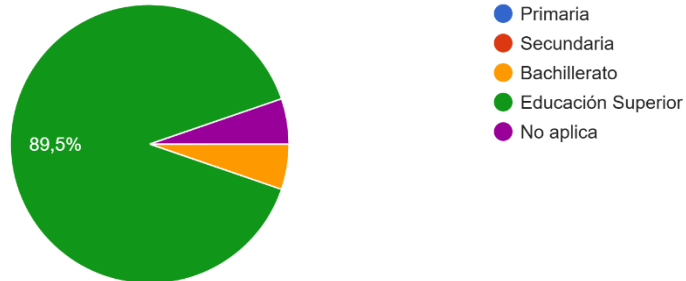
El 42,1% asistió a una institución privada y un pequeño porcentaje a una fiscal-mixta, lo que refleja que hay una diversidad en los tipos de instituciones, pero con mayor presencia de lo público.

**Figura 4**

*Pregunta 4*

¿En qué nivel educativo se encuentra actualmente o dejó sus estudios?

19 respuestas



La mayoría, es decir, el 89,5%, señaló que se encuentra o estuvo en educación superior, lo cual muestra que gran parte de los encuestados accedió a estudios universitarios o tecnológicos.

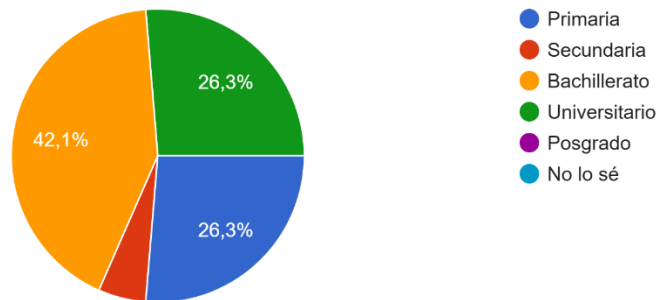
Por otro lado, un pequeño porcentaje pertenece a bachillerato y educación secundaria, lo que representa una minoría dentro de la muestra. Es importante destacar que ninguna persona indicó estar en primaria, lo cual era previsible dado el rango de edad observado previamente.

**Figura 5**

*Pregunta 5*

Cuál es el nivel educativo más alto alcanzado por sus padres o tutores?

19 respuestas



En este gráfico se ve hasta qué nivel estudiaron los padres o tutores. La mayoría terminó el bachillerato con un 42,1%, lo que muestra que muchos padres tienen un nivel medio de educación.

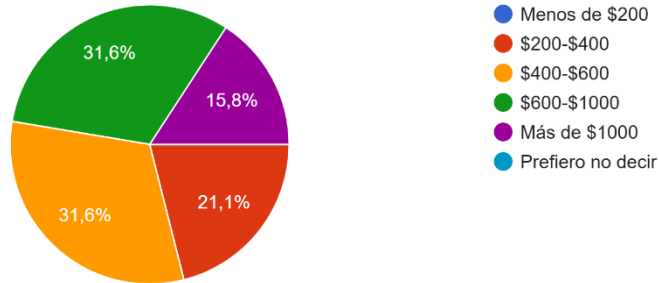
Luego, el 26,3% llegó a la universidad y otro 26,3% solo completó la primaria. Esto es importante porque el nivel de estudios de los padres puede influir en el apoyo que dan a sus hijos para seguir estudiando.

**Figura 6**

*Pregunta 6*

Cuál es el ingreso mensual aproximado de su hogar?

19 respuestas



En este gráfico se muestra el ingreso mensual aproximado de los hogares. El 31,6% de los encuestados indicó que sus hogares reciben entre \$400 y \$600, y otro 31,6% mencionó que ganan entre \$600 y \$1000.

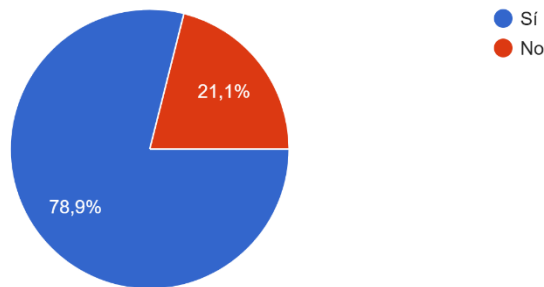
Un 21,1% dijo que su ingreso es entre \$200 y \$400, y el 15,8% informó que gana más de \$1000 al mes. Esto refleja que la mayoría de familias tiene ingresos medios o bajos, lo cual puede influir en las decisiones sobre continuar o no los estudios.

**Figura 7**

*Pregunta 7*

¿Trabaja actualmente o ha trabajado mientras estudiaba?

19 respuestas



Este gráfico muestra que el 78,9% de los encuestados trabaja o trabajó mientras estudiaba, lo cual refleja que la mayoría tuvo que combinar sus estudios con actividades laborales.

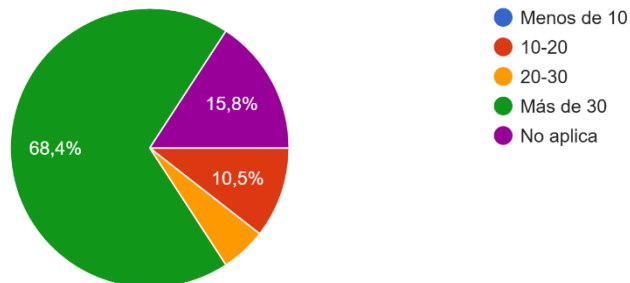
Por otro lado, el 21,1% indicó que nunca trabajó durante sus estudios, lo que podría significar que contaron con un mayor apoyo económico o condiciones más favorables para dedicarse solo a estudiar.

**Figura 8**

*Pregunta 8*

Si trabaja, ¿cuántas horas semanales dedica al trabajo?

19 respuestas



En este gráfico se observa que el 68,4% trabaja más de 30 horas a la semana, lo que indica que la mayoría dedica casi una jornada completa al trabajo mientras estudia.

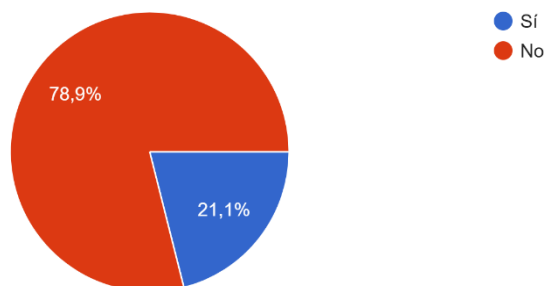
El 10,5% trabaja entre 10 y 20 horas, y un 15,8% no aplica porque no trabaja. Esto muestra que combinar trabajo y estudio es común entre los encuestados y puede afectar su rendimiento académico.

**Figura 9**

*Pregunta 9*

¿Recibe algún tipo de apoyo económico o beca para estudiar?

19 respuestas



Este gráfico muestra que el 78,9% de los encuestados no recibe apoyo económico ni beca para estudiar, lo que significa que deben cubrir sus gastos educativos por cuenta propia o con ayuda familiar.

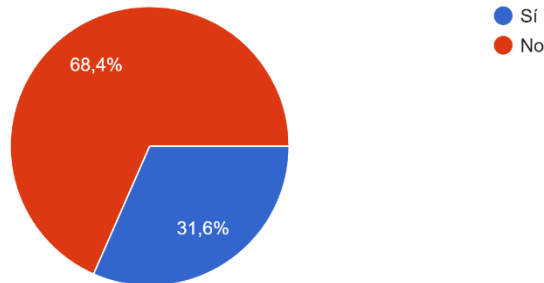
Solo el 21,1% cuenta con algún tipo de beca o apoyo económico, lo que refleja que el acceso a estos beneficios es limitado y podría influir en el riesgo de abandonar los estudios.

**Figura 10**

*Pregunta 10*

¿Ha considerado dejar sus estudios en algún momento?

19 respuestas



Este gráfico muestra que el 31,6% de los encuestados ha pensado en dejar sus estudios, lo que refleja que una parte significativa ha tenido dificultades o dudas sobre continuar.

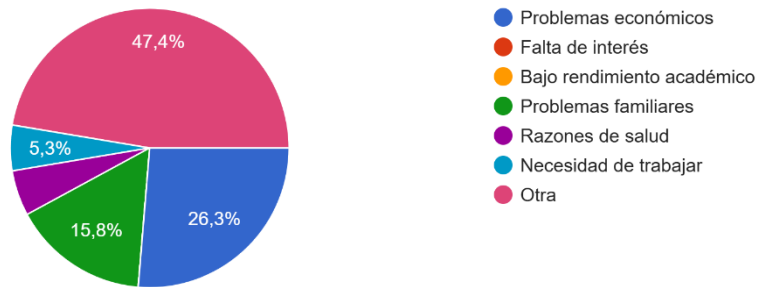
Sin embargo, el 68,4% indicó que nunca ha considerado abandonar sus estudios, lo cual es positivo, aunque es importante analizar qué factores ayudan a que sigan adelante.

**Figura 11**

*Pregunta 11*

Si dejó los estudios, ¿cuál fue el motivo principal?

19 respuestas



En este gráfico se observa que el 47,4% de los encuestados señaló "otra razón" como el motivo principal para dejar sus estudios, lo que indica que hay causas diversas que no siempre se contemplan en las opciones más comunes.

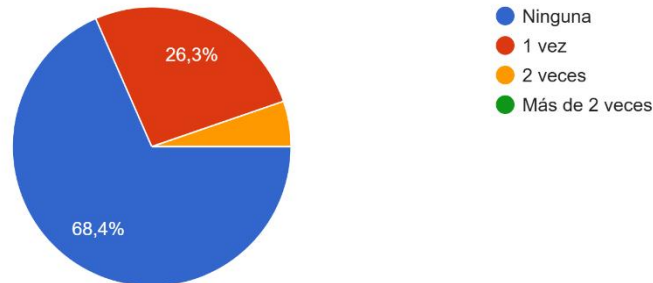
El 26,3% mencionó problemas económicos, mientras que el 15,8% indicó problemas familiares. Esto muestra que los factores económicos y familiares siguen siendo causas importantes de escolaridad inconclusa.

**Figura 12**

*Pregunta 12*

¿Cuántas veces ha repetido un curso o grado escolar?

19 respuestas



Este gráfico muestra que el 68,4% de los encuestados nunca ha repetido un curso o grado escolar, lo cual es un dato positivo que refleja cierta estabilidad académica.

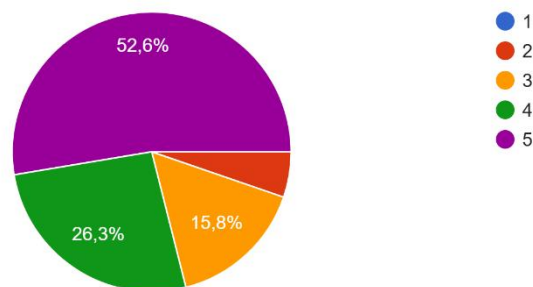
Sin embargo, el 26,3% ha repetido al menos una vez, y un pequeño porcentaje ha repetido dos veces. Esto sugiere que algunos estudiantes han enfrentado dificultades en su proceso educativo.

**Figura 13**

*Pregunta 13*

¿Se siente motivado para continuar con sus estudios? (Escala de 1 a 5, donde 1 es nada motivado y 5 es muy motivado)

19 respuestas



En este gráfico se observa que el 52,6% de los encuestados se siente muy motivado para continuar con sus estudios, lo cual es un dato positivo que refleja interés y compromiso.

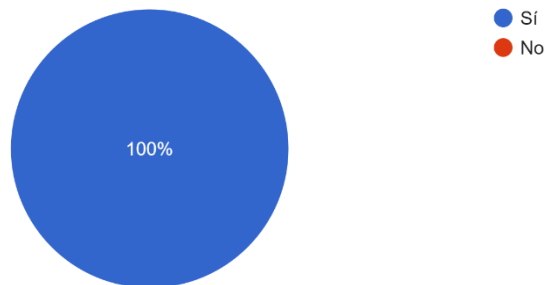
Por otro lado, el 26,3% tiene una motivación media, mientras que el 15,8% reporta una baja motivación. Esto muestra que, aunque la mayoría está motivada, hay un grupo que podría necesitar más apoyo para seguir adelante.

**Figura 14**

*Pregunta 14*

¿Cuenta con acceso a internet y dispositivos electrónicos para estudiar?

19 respuestas



En este gráfico se muestra que el 100% de los encuestados tiene acceso a internet y dispositivos electrónicos para estudiar, lo cual es una ventaja importante para el desarrollo académico.

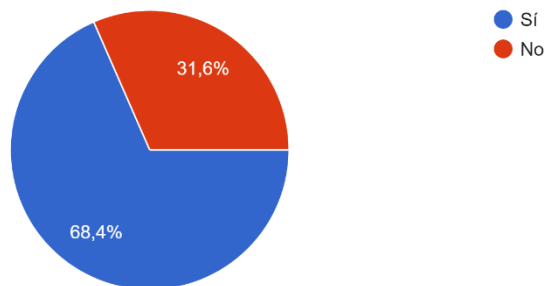
Esto indica que, al menos en el grupo encuestado, el acceso tecnológico no es una barrera directa para la continuidad de sus estudios, aunque otros factores como económicos o familiares sí podrían influir.

**Figura 15**

*Pregunta 15*

¿Ha tomado clases virtuales o a distancia en los últimos dos años?

19 respuestas



Este gráfico muestra que el 68,4% de los encuestados ha tomado clases virtuales o a distancia en los últimos dos años, lo cual refleja el impacto que ha tenido la modalidad virtual en su formación.

El 31,6% no ha tenido esa experiencia, lo que podría deberse a su tipo de institución o a la modalidad presencial que han mantenido. Esto es importante porque las clases virtuales pueden influir en la motivación y el rendimiento.

## **CAPÍTULO II: PROPUESTA**

### **1.1. Fundamentos teóricos aplicados**

#### **Definición de Escolaridad Inconclusa**

Se entiende que la escolaridad inconclusa es la interrupción de la formación académica antes de alcanzar a un grado educativo determinado (educación básica, media y/o superior). Para Vélez (2022), la escolaridad inconclusa identifica una práctica frecuente de modo que va relacionada a la discontinuidad de los estudiantes en su proceso educativo que es consecuencia de diversas intervenciones que se encuentran enmarcadas entre lo familiar, lo socioeconómico y lo educativo; es en la interrupción de la escolaridad donde hay un fenómeno en el curso de los estudiantes que puede estar relacionado con la escolaridad inconclusa especialmente en el contexto ecuatoriano, la escolaridad inconclusa configura también no solo un problema educativo sino que también limita la formación académica de los jóvenes, esto también limita el desarrollo económico y social del país. De forma proporcional la escolaridad inconclusa que sucede dentro de la teoría puede explicarse a partir de teorías como el Capital Cultural de Bourdieu, aplicado a cómo el nivel educativo por parte de las familias y su contexto sociocultural tiene influencia directa en el rendimiento y en la desistencia que tienen los estudiantes para permanecer en el sistema educativo (Bourdieu, 1997). Esta base teórica es relevante para poder explicitar la capacidad que tienen las condiciones sociales y económicas en el proceso de la deserción escolar.

#### **Concepto de Machine Learning y su Aplicación**

El Machine Learning, o aprendizaje automático, es una rama de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos, sin necesidad de ser programados de manera explícita. Según Forero y Bennisar (2024) un sistema de Machine Learning mejora su desempeño en una tarea determinada conforme recibe más datos y genera patrones predictivos. En el contexto de esta propuesta, el uso de Machine Learning permitirá analizar datos históricos de los estudiantes registrados en el Ministerio de Educación para identificar patrones asociados a la escolaridad inconclusa y predecir el riesgo de abandono.

El fundamento teórico de aprendizaje supervisado, descrito por los autores, es esencial en esta propuesta. En este enfoque, la ejecución del modelo recibe un conjunto de datos etiquetados (estudiantes que abandonaron y estudiantes que continuaron) y aprende a reconocer las características que diferencian a ambos grupos. Posteriormente, es capaz de

predecir el riesgo de abandono para nuevos estudiantes. Este enfoque teórico es clave para la construcción del modelo predictivo.

### **Big Data Educativa y su importancia**

El concepto de Big Data Educativa se refiere a la recogida, almacenamiento y análisis de grandes volúmenes de datos que son generados dentro del sistema educativo. El análisis de datos masivos en educación, según Guña (2023), significa que se despliegan procedimientos para encontrar patrones ocultos en los datos, deducir comportamientos y mejorar la toma de decisiones que afectan las instituciones educativas. En lo que concierne a esta propuesta, la Big Data permitirá trabajar con bases de datos extensas del MINEDUC mediante la recogida, la exploración y el análisis de variables, sociodemográficas, académicas y familiares que influyen directamente en la permanencia escolar.

Teóricamente, el análisis de Big Data Educativa está sustentado también en el Enfoque Predictivo, como son descritos por Heredia (Heredia, 2020), el análisis de grandes volúmenes de datos históricos es la clave para anticipar eventos, por ejemplo, una de estas predicciones podría ser la del riesgo de deserción escolar; este enfoque permite proporcionar la base metodológica para el futuro desarrollo del modelo de Machine Learning.

### **Teoría de la Intervención Temprana**

Un concepto importante vinculado con la propuesta es la intervención temprana. Se basa en la detección, de forma precoz, de estudiantes en condiciones de riesgo para poder potenciar estrategias de apoyo antes de que se produzca el abandono del sistema. Balseca et al., (2024) señala que las intervenciones precoces tienen un impacto superior a las acciones correctivas de forma posterior, ya que se enfocan en los factores de riesgo antes de que estos se consoliden; este enfoque teórico evidencia la importancia de contar con un modelo predictivo automático, con el objetivo de poder situar y, por tanto, poder diseñar acciones específicas de acompañamiento de los estudiantes en riesgo.

### **Relación entre Tecnología y Educación**

En última instancia, Tumbaco (2022) indica que las herramientas tecnológicas no solo abren la vía al conocimiento, sino que permiten transformar los procesos de gestión y análisis educativos, por lo que recurrir a Machine Learning y Big Data no solo es posible en el ámbito

educativo, sino necesaria para modernizar los procesos de monitoreo y evaluación educativa en Ecuador.

## **1.2. Descripción de la propuesta**

La propuesta se centra en ejecutar un modelo predictivo de Machine Learning que se podrá utilizar para explicar la escolaridad inconclusa de los estudiantes que están registrados en el Ministerio de Educación del Ecuador. Este modelo aprovecha el acceso a bases de datos públicas, como las que se ofrecen en el portal de Datos Abiertos del MINEDUC, para acceder a información histórica de estudiantes, instituciones, niveles, tasas de abandono, promoción y no promoción y variables sociodemográficas.

El modelo predictivo se alimentará principalmente de la base oficial de instituciones educativas, la cual contiene variables que son representativas de la escolaridad inconclusa como: zona, provincia, cantón, parroquia, nombre de la institución, tipo de sostenimiento (fiscal, fiscomisional, municipal o privado), modalidad educativa (presencial, semipresencial, a distancia), acceso a internet, número de docentes, estudiantes con y sin discapacidad, nivel educativo, entre otros. Además, se incorporarán indicadores directamente relacionados con el abandono escolar de esta manera el modelo se entrena basándose en datos de los estudiantes reales y actualizados.

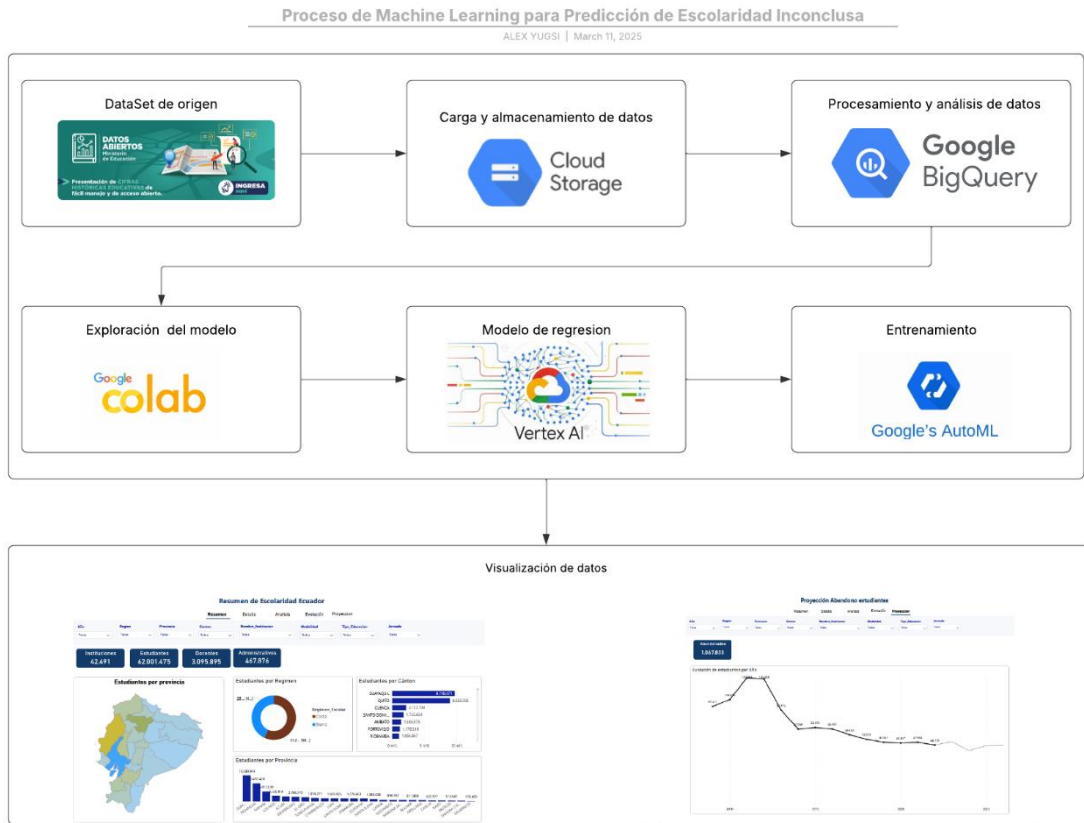
Como producto final se aplicará un modelo analítico de Machine Learning que permita analizar la información del abandono y predecir los estudiantes con mayor riesgo de abandono escolar. Este modelo estará construido en base a algoritmos de regresión para detectar patrones de abandono escolar. Estará diseñado para complementarse con el sistema de monitoreo educativo del MINEDUC que les responderá a las autoridades alertándoles anticipadamente para que puedan responder de manera anticipada y prevenir el abandono escolar.

De esta manera, el modelo se convierte en una herramienta tecnológica que permite reducir la escolaridad inconclusa y mejorar las estrategias para prevenir el abandono de la escolaridad en Ecuador.

## a. Estructura general

Figura 16

Estructura General de la Propuesta



Nota. Se muestra las etapas clave del modelo de Machine Learning para analizar la escolaridad inconclusa.

## b. Explicación del aporte

El dataset contiene información detallada sobre instituciones educativas, docentes y estudiantes

Figura 17

Estructura de datos del ministerio de educación

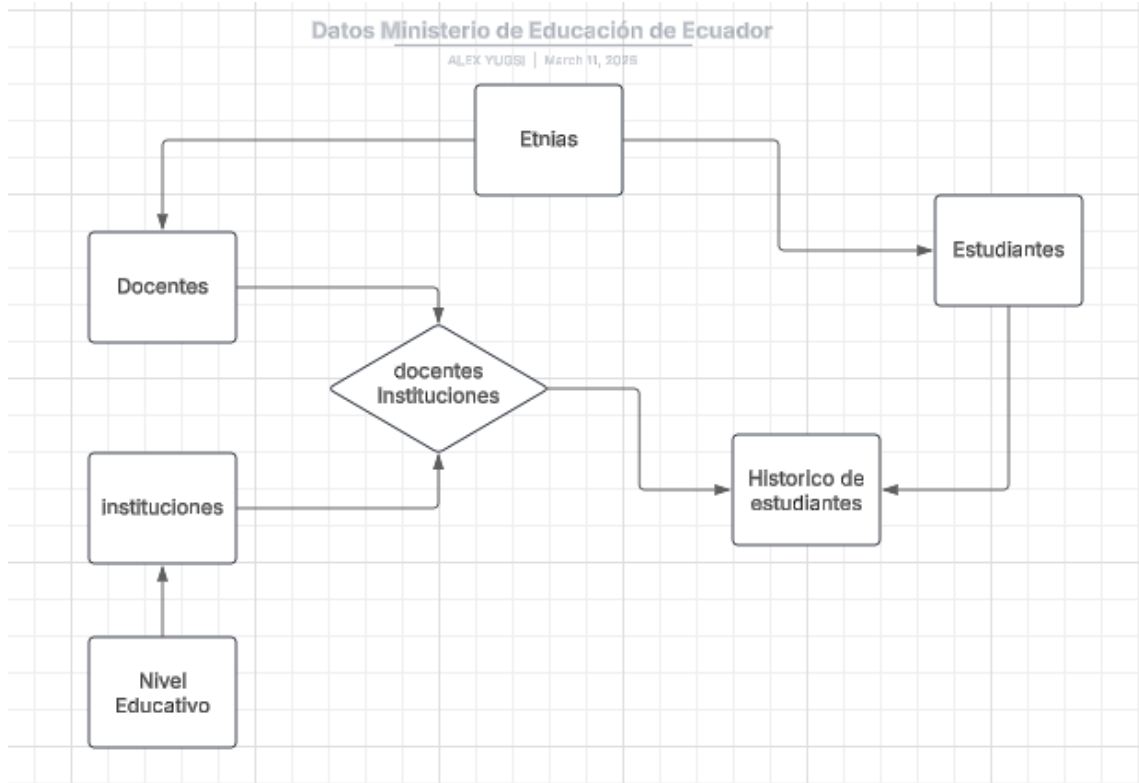


Figura 18

Datos históricos de estudiantes.

Consulta sin título

`SELECT * FROM 'proyecto-mineduc-451102.educacion_datos.vista_registro_historico' LIMIT 1000`

Ubicación de procesamiento: US

Presiona Alt+F1 para ver las opciones de accesibilidad.

Resultados de la consulta

| INFORMACIÓN DEL TRABAJO |    | RESULTADOS   | GRÁFICO            | JSON                  | DETALLES DE LA EJECUCIÓN |               | GRÁFICO DE EJECUCIÓN |          |
|-------------------------|----|--------------|--------------------|-----------------------|--------------------------|---------------|----------------------|----------|
| Fila                    | to | Nombramiento | Otro_Tipo_Relacion | Total_Estudiantes_Fil | Promovidos               | No_Promovidos | N_Abandono           | Abandono |
| 1                       | 9  | 1            | 0                  | 47                    | 47                       | 0             | 0                    | 0        |
| 2                       | 12 | 5            | 0                  | 264                   | 219                      | 30            | 15                   | 1        |
| 3                       | 1  | 0            | 0                  | 24                    | 17                       | 5             | 2                    | 1        |
| 4                       | 2  | 0            | 3                  | 273                   | 256                      | 2             | 15                   | 1        |
| 5                       | 0  | 1            | 4                  | 0                     | 0                        | 0             | 0                    | 0        |
| 6                       | 2  | 3            | 0                  | 130                   | 109                      | 9             | 12                   | 1        |
| 7                       | 1  | 0            | 0                  | 17                    | 12                       | 3             | 2                    | 1        |
| 8                       | 1  | 1            | 0                  | 55                    | 49                       | 6             | 0                    | 0        |
| 9                       | 4  | 1            | 0                  | 78                    | 77                       | 1             | 0                    | 0        |
| 10                      | 1  | 3            | 0                  | 89                    | 76                       | 4             | 9                    | 1        |

Resultados por página: 50 1 - 50 de 1000

Se observa que una de las columnas de interés es "Abandono", la cual indica la cantidad de estudiantes que han abandonado sus estudios. Dado que el objetivo del modelo es analizar la escolaridad inconclusa, esta variable es clave en la selección del algoritmo.

En este caso, dado que tenemos una variable de salida clara (Abandono), el problema es de aprendizaje supervisado.

### **Algoritmo Supervisado**

**Disponibilidad de etiquetas.** La variable "Abandono" proporciona datos históricos sobre la cantidad de estudiantes que han abandonado la educación en cada institución, lo que nos permite entrenar un modelo con datos conocidos.

**Objetivo de predicción.** Queremos predecir si un estudiante o institución tiene alta probabilidad de presentar abandono escolar en función de factores socioeconómicos, geográficos y de infraestructura. Esto requiere un modelo que relacione características de entrada con una salida específica.

**Estructura del dataset.** Contamos con múltiples características (zona, nivel educativo, acceso a internet, tipo de contratación de docentes, entre otros) que pueden servir como variables predictoras. En problemas de predicción como este, los modelos supervisados suelen ofrecer mejor precisión.

**Facilidad de validación.** Dado que tenemos datos históricos etiquetados, podemos evaluar la precisión del modelo comparando sus predicciones con los datos reales de abandono.

#### ***Adquisición y Exploración de Datos***

**Cargar y examinar el dataset.** Revisamos las columnas, tipos de datos y si hay valores faltantes.

**Limpieza de datos.** Eliminamos valores nulos o inconsistencias y realizaremos transformaciones necesarias.

**Análisis exploratorio de datos (EDA).** Identificamos patrones, correlaciones y distribución de variables.

El dataset usado para el proyecto se encuentra en una hoja de Excel es por eso que se define como datos estructurados

Los datos estructurados son información organizada y formateada que sigue un modelo predefinido. Se almacenan en bases de datos y en almacenes de datos.

## **Características**

- Son fáciles de procesar por las personas y los algoritmos de Machine Learning.
- Permiten tomar decisiones más informadas y fundamentadas.
- Son cuantitativos, es decir, se utilizan para expresar volúmenes, cantidades o un rango de valores.
- Se pueden ingresar en hojas de cálculo o bases de datos relacionales

## **Proceso**

Para el desarrollo del presente proyecto, se utilizaron herramientas de Google Cloud con el objetivo de analizar y predecir la escolaridad inconclusa en estudiantes. El proceso seguido fue el siguiente

**Carga y almacenamiento de datos.** Los datos históricos fueron almacenados en Google Cloud Storage (GCS), permitiendo una gestión escalable y segura de la información.

**Procesamiento y análisis de datos.** Se utilizó Google BigQuery para realizar consultas SQL sobre los datos almacenados, facilitando la limpieza, transformación y análisis exploratorio.

**Exploración y desarrollo del modelo.** Con Google Colab, se realizó la experimentación inicial del modelo de Machine Learning, evaluando diferentes algoritmos y métricas de rendimiento.

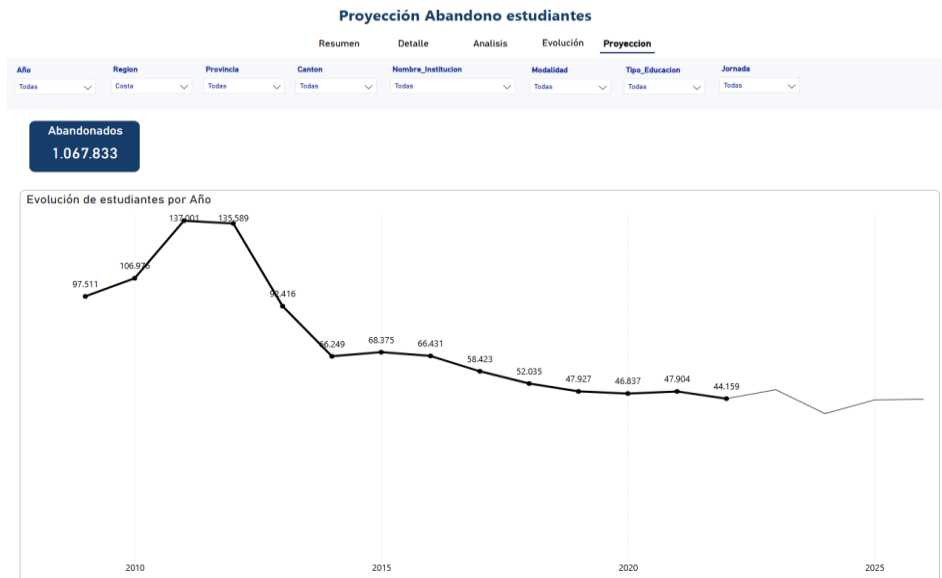
**Entrenamiento y despliegue del modelo.** A través de Google Vertex AI y AutoML, se entrenó y optimizó el modelo de predicción de abandono escolar

**Visualización de resultados.** Finalmente, los resultados fueron integrados en Power BI, permitiendo la creación de dashboards interactivos para el análisis de tendencias y la toma de decisiones informadas.

## Resultado final

Figura 19

Proyección abandono de estudiantes



## Modelo aplicado

En este proyecto se aplicaron Regresión Lineal y Regresión Multivariable con el objetivo de predecir y analizar la escolaridad inconclusa en estudiantes.

Según Montgomery et al. (2021), la Regresión Lineal ha sido utilizada con frecuencia en investigaciones relacionadas con el ámbito educativo, ya que permite establecer la relación entre variables como el rendimiento académico y la probabilidad de deserción escolar. Su enfoque es útil para comprender cómo un solo factor puede influir en un resultado, facilitando el análisis de tendencias en poblaciones estudiantiles.

Por otro lado, la Regresión Multivariable resulta más adecuada cuando es necesario considerar múltiples factores que pueden intervenir en un fenómeno. James et al. (2021) explican que este tipo de análisis es clave en estudios educativos porque permite evaluar simultáneamente la interacción de variables.

García y López (2020) destacan que el uso de estos modelos no solo facilita la identificación de factores de riesgo en la deserción escolar, sino que también contribuye a diseñar estrategias para mejorar la permanencia de los estudiantes.

### Uso de Regresión Lineal

**Motivo.** Se utilizó para analizar la relación entre una única variable independiente y la variable objetivo, es decir, predecir la tasa de abandono escolar en función de un solo factor clave.

- Permite entender cómo varía el abandono escolar en función de una variable (ejemplo: nivel socioeconómico, acceso a internet, número de docentes por estudiante).
- Se obtiene una ecuación lineal simple que facilita la interpretación y toma de decisiones.
- Es útil para realizar análisis exploratorios iniciales antes de implementar modelos más complejos.

### Uso de Regresión Multivariable

**Motivo.** Se utilizó cuando se identificó que múltiples factores influían en la escolaridad inconclusa, permitiendo modelar relaciones más complejas entre varias variables.

Permite analizar el efecto combinado de múltiples factores en la predicción del abandono escolar.

Se obtiene una ecuación en la forma:

$$\text{Abandono} = \beta_0 + \beta_1(X_1) + \beta_2(X_2) + \dots + \beta_n(X_n) + \epsilon$$

Donde cada  $X_i$  representa una variable independiente como zona geográfica, tipo de institución, acceso a internet, número de promovidos, entre otros.

Mejora la precisión de la predicción al incluir múltiples determinantes del abandono escolar.

Es compatible con técnicas de regularización (Lasso y Ridge) para evitar el sobreajuste y mejorar la generalización.

### Para el proyecto se usa algoritmos supervisados específicamente Algoritmos de regresión

Los algoritmos supervisados son un tipo de aprendizaje automático que se basa en conjuntos de datos etiquetados para predecir resultados. Son una de las técnicas de aprendizaje automático más usadas.

Cómo funcionan

Se entrena un algoritmo con un conjunto de datos de entrenamiento que incluye datos de entrada etiquetados

El algoritmo aprende a establecer relaciones entre las características y los datos de salida

El algoritmo puede predecir los valores de respuesta para un nuevo conjunto de datos

Tipos de algoritmos supervisados

- Algoritmos de regresión
- Algoritmos de clasificación
- Árboles de decisión

### Herramientas de Google Cloud para Machine Learning

El uso de herramientas como BigQuery, Google Cloud Storage, Vertex AI y AutoML en lugar de tecnologías más tradicionales como Apache Spark, Hive y Pig se fundamenta en varios factores clave:

**Tabla 1**

*Elección de Herramientas de Google Cloud para Machine Learning*

| <b>Criterio</b>         | <b>Google Cloud (BigQuery, Vertex AI, AutoML)</b>                             | <b>Apache Spark, Hive, Pig</b>  |
|-------------------------|---|---|
| <b>Escalabilidad</b>    | Alta (100% en la nube, sin gestión de infraestructura) (Google Cloud, 2024).  | Requiere clústeres y configuración manual (AWS, 2024)                     |
| <b>Facilidad de uso</b> | Alta (interfaces gráficas, no requiere código avanzado) (Google Cloud, 2024). | Baja (requiere conocimientos en Scala, Python, SQL avanzado) (AWS, 2024). |
| <b>Automatización</b>   | Sí (AutoML ajusta hiperparámetros automáticamente) (Google Cloud, 2024).      | No (debe programarse manualmente) (AWS, 2024).                            |
| <b>Costo operativo</b>  | Menor (servicios serverless y pagos por uso) (Google Cloud, 2024).            | Mayor (clústeres requieren administración) (AWS, 2024).                   |
| <b>Integración</b>      | Nativa con herramientas de Google (Google Cloud, 2024).                       | Requiere configuraciones adicionales (AWS, 2024).                         |
| <b>Seguridad</b>        | Encriptación automática y cumplimiento de normativas (Google Cloud, 2024).    | Configuración manual de seguridad (AWS, 2024).                            |

Google Cloud (BigQuery, Vertex AI, AutoML) es la mejor opción cuando se busca escalabilidad, facilidad de uso, integración y seguridad sin administrar infraestructura compleja

## Costo de las herramientas.

Tabla 2

Costo – Beneficio Google Cloud

| Herramienta                 | Plan Básico /<br>Gratuito   | Plan Pago   | Características   |
|-----------------------------|---|---|---|
| <b>Google BigQuery</b>      | 10 GB de almacenamiento gratis (Google Cloud, 2024).              | \$5 por TB de consulta; \$0.02 por GB al mes de almacenamiento adicional (Google Cloud, 2024).                                      | Consulta de datos con SQL; almacenamiento escalable (Google Cloud, 2024).   |
| <b>Google Cloud Storage</b> | No aplica   | \$0.026 por GB al mes (Clase Estándar); \$0.01 por GB al mes (Clase Nearline) (Google Cloud, 2024).                                 | Almacenamiento en la nube con diferentes clases de acceso (Google Cloud, 2024).   |
| <b>Google Colab</b>         | Acceso gratuito con recursos limitados (Google Cloud, 2024).      | \$9.99/mes (Colab Pro); \$49.99/mes (Colab Pro+) (Google Cloud, 2024).  | Ejecución de código en GPU/TPU; mayor tiempo de ejecución (Google Cloud, 2024).   |
| <b>Vertex AI</b>            | Créditos promocionales para nuevos usuarios (Google Cloud, 2024). | Costos variables según recursos utilizados (Google Cloud, 2024).  | Plataforma unificada de Machine Learning; integración con otros servicios de Google Cloud (Google Cloud, 2024).                                   |
| <b>AutoML</b>               | Créditos promocionales para nuevos usuarios (Google Cloud, 2024). | Costos variables según el tipo de modelo y recursos utilizados (Google Cloud, 2024).  | Creación automática de modelos de Machine Learning personalizados; no requiere experiencia previa en aprendizaje automático (Google Cloud, 2024). |
| <b>Power BI</b>             | No aplica   | \$9.99 por usuario al mes (Pro); desde \$20 por usuario al mes o \$4,995 por capacidad dedicada al mes (Premium) (Microsoft, 2024). | Creación de dashboards interactivos; integración con múltiples fuentes de datos (Microsoft, 2024).  |

### c. Técnicas

Para la elaboración de esta propuesta, se llevaron a cabo diversas estrategias y técnicas de Data Science y Machine Learning con el objetivo de asegurar que el modelo predictivo tenga un correcto funcionamiento y aporte información veraz en la toma de decisiones del Ministerio de Educación del Ecuador, estas estrategias abarcan los procesos clave que se van desde la recolección de datos hasta la interpretación de los resultados que se obtengan.

**Recolección de datos abiertos.** Se recuperó información desde el portal de Datos Abiertos del MINEDUC, que incluye registros históricos de estudiantes, niveles educativos, tasas de promoción y deserción, tipo de institución, modalidad educativa y factores sociodemográficos.

**Preprocesamiento.** Se implementaron diversas técnicas de limpieza de datos como eliminación de registros duplicados, tratamiento de valores vacíos y codificación de variables categóricas para así poder ser procesados por el modelo propuesto.

**Selección de variables.** Se seleccionaron las variables más relevantes para la predicción de la escolaridad inconclusa, por ejemplo, tipo de institución, sostenimiento, acceso a internet, modalidad de estudio, jornada, nivel educativo. Estas variables de las definido en base en la aplicación de un análisis de componentes principales (CPA).

**Visualización de resultados.** Se generaron gráficos interpretativos con el fin de facilitar la comprensión y la lectura de los patrones que se identifican, de tal forma que las autoridades educativas puedan tomar decisiones adecuadas. Se usaron las variables obtenidas del análisis CPA.

Estas estrategias garantizan construir un modelo sólido, confiable y que pueda ser adaptado a la realidad educativa ecuatoriana y que pueda ser una herramienta muy efectiva y con el potencial de ayudar a reducir la deserción escolar mediante la identificación anticipada del riesgo que presentan los estudiantes.

#### 1.3. Validación de la propuesta

Para la construcción del modelo de Machine Learning que analiza la escolaridad inconclusa en estudiantes del Ministerio de Educación, se implementaron diversas estrategias y técnicas basadas en Data Science, garantizando la calidad de los datos y la precisión de las predicciones.

Los resultados de esta validación, junto con la retroalimentación proporcionada por los especialistas, se encuentran detallados en el **Anexo 3 y Anexo 4**

Se implementaron herramientas de visualización de datos para facilitar la interpretación de patrones y tendencias. Estos reportes permiten a las autoridades del MINEDUC identificar zonas de alto riesgo y tomar medidas preventivas.

Estas estrategias garantizan que el modelo sea confiable y adaptable a la realidad educativa ecuatoriana, proporcionando una herramienta efectiva para reducir la deserción escolar mediante la detección temprana de estudiantes en riesgo.

#### 1.4. Matriz de articulación de la propuesta

En la presente matriz se sintetiza la articulación del producto realizado con los sustentos teóricos, metodológicos, estratégicos-técnicos y tecnológicos empleados.

**Tabla 3**

*Matriz de articulación*

| EJES O PARTES PRINCIPALES        | SUSTENTO TEÓRICO  | SUSTENTO METODOLÓGICO  | ESTRATEGIAS / TÉCNICAS   | DESCRIPCIÓN DE RESULTADOS  | INSTRUMENTOS APLICADOS                          |
|----------------------------------|---|--|--|--|---|
| <b>Recolección de datos</b>      | Se basa en el uso de datos abiertos proporcionados por el Ministerio de Educación (MINEDUC), fundamentado en la importancia del análisis de datos en la educación (Benalcázar, 2024). | Se recopiló información de bases de datos oficiales del MINEDUC, utilizando un enfoque cuantitativo para garantizar precisión en la muestra. | Extracción de datos abiertos, consolidación y almacenamiento en un entorno de procesamiento estructurado.            | Se obtuvo un conjunto de datos amplio y representativo con información sobre estudiantes, niveles educativos, instituciones y tasas de abandono. | Portal de Datos Abiertos del MINEDUC            |
| <b>Preprocesamiento de datos</b> | Fundamentado en técnicas de Data Cleaning y Transformación de Datos (Jaramillo, 2022) para mejorar la calidad de los  | Fundamentado en técnicas de Data Cleaning y Transformación de Datos (Jaramillo, 2022) para mejorar la calidad                                | Técnicas de limpieza de datos, tratamiento de valores perdidos y codificación de variables cualitativas a numéricas. | Se obtuvo un dataset listo para ser procesado por algoritmos de Machine Learning, con información estructurada y sin errores.                    | Lenguajes de programación como Python, BigQuery |

|  |  |  |   |   |  |
|--|--|--|---|---|--|
|  | datos antes del análisis.  | de los datos antes del análisis.   |   |   |  |
| <b>Entrenamiento del modelo</b>              | Basado en los principios de Machine Learning supervisado y su aplicación en educación para la detección de patrones predictivos (Freire, 2023).                  | Se implementaron modelos de regresión para predecir la escolaridad inconclusa, evaluando distintas técnicas y métricas de rendimiento. | Uso de algoritmos de regresión  | Se seleccionó el modelo con mejor desempeño, capaz de predecir el abandono escolar con alta precisión.                      | Herramientas de Google como Vertex IA, Auto ML       |
| <b>Generación de alertas y visualización</b> | Se apoya en la importancia de los Sistemas de Soporte a Decisiones (DSS) en educación (Tumbaco, 2022), permitiendo acciones tempranas ante riesgos de deserción. | Se crearon reportes con predicciones del modelo, visualización de datos  | Dashboards interactivos, gráficas y reportes con información clara para el MINEDUC. | Se logró ejecutar un modelo predictivo que permite detectar estudiantes en riesgo, facilitando estrategias de intervención. | Herramientas de visualización de datos como Power BI |

Fuente: Elaboración propia

## CONCLUSIONES

Las conclusiones del trabajo en función de los objetivos propuestos son las siguientes:

Se ha establecido un sólido marco teórico que acompaña la aplicación de Machine Learning como herramienta predictiva de la deserción escolar, ya que mediante la revisión de la literatura fue posible comprobar que la combinación de datos relativos a la educación y modelos predictivos resulta capaz de detectar patrones de deserción anticipadamente.

Se utilizaron los datos abiertos del MINEDUC para conseguir la identificación de factores que inciden en la deserción escolar, ya que a través de él se obtuvo información relevante, como el tipo de institución, la modalidad de estudio, la accesibilidad a internet, el nivel socioeconómico y la jornada escolar. Gracias a estos resultados fue posible la elección de las variables que se utilizaran para entrenar el modelo predictivo.

Se generó un dashboard que permite interpretar los datos de forma didáctica y en consecuencia tomar decisiones de prevención, emitiendo alertas tempranas para con los estudiantes.

Esta investigación ha permitido mostrar que los resultados de aplicar Machine Learning en la educación constituyen herramientas para anticipar la deserción escolar y mejorar el resultado de las acciones de intervención. El modelo decidido ha proporcionado una respuesta innovadora y por tanto se intenta integrar y aplicar en el sistema educativo ecuatoriano, con la finalidad de reducir la escolaridad inconclusa, así como de mejorar el rendimiento en la tasa de retención.

## RECOMENDACIONES

Se sugiere ampliar el conocimiento sobre otros modelos de persona inteligente y su repercusión en la disminución de la deserción escolar, así como futuras investigaciones que exploren modelos híbridos con datos de aprendizaje supervisado y no supervisado para aumentar la efectividad de las predicciones. También sería conveniente añadir otras variables como factores emocionales, apoyo familiar, salud mental, entorno comunitario, etc., de forma tal que se permita la detección de riesgos. Sumando a los datos: el uso de datos cualitativos podría servir de complemento para enriquecer la información usada para entender el problema y la situación.

Además, sería recomendable realizar una evaluación continua del modelo, sistemáticamente actualizando los datos y ajustando los algoritmos en función de las tendencias educativas y de los cambios en las políticas del MINEDUC, para así garantizar la eficacia y la exactitud del sistema en los pronósticos de abandono escolar.

Se aconseja al Ministerio de Educación que utilice las predicciones del mismo para diseñar estrategias de intervención temprana, como por ejemplo programar tutorías, ofrecer apoyo psicológico o becas para los estudiantes que el modelo evidencie con mayor riesgo de deserción.

Es necesario que se divulguen y se diseminan los resultados de la investigación en seminarios, congresos de educación, en publicaciones científicas etc., para propiciar la transición hacia la utilización de herramientas tecnológicas para tomar decisiones educativas.

## BIBLIOGRAFÍA

- AWS. (2024). Diferencias entre Hadoop y Spark. Recuperado el 10 de marzo de 2025, de <https://aws.amazon.com/es/compare/the-difference-between-hadoop-vs-spark/>
- Benalcázar, E. V. (2024). *Evaluación del aprendizaje en estudiantes con escolaridad inconclusa mediante recursos digitales*. *MQRInvestigar*, 8(1), 2859-2878.  
<http://www.investigarmqr.com/ojs/index.php/mqr/article/download/1032/3874>
- Freire, H. (2023). *Desarrollo del pensamiento lógico matemático en estudiantes de escolaridad inconclusa (Master's thesis)*.  
<https://dspace.ups.edu.ec/bitstream/123456789/24161/1/UPS-CT010299.pdf>
- García, J., & López, M. (2020). Modelos estadísticos aplicados a la educación: Predicción y análisis de datos. Editorial Académica.
- Google. (2023). Google Colaboratory. Recuperado el 10 de marzo de 2025, de <https://colab.research.google.com/>
- Google Cloud. (2024). Google Cloud Dataproc: Managed Spark and Hadoop service. Recuperado el 10 de marzo de 2025, de <https://cloud.google.com/dataproc>
- Google Cloud. (2024). BigQuery Pricing. Recuperado el 10 de marzo de 2025, de <https://cloud.google.com/bigquery/pricing>
- Google Cloud. (2024). Google Colab Pricing., de <https://colab.research.google.com/signup>
- Google Cloud. (2024). Vertex AI Pricing. Recuperado el 10 de marzo de 2025, de <https://cloud.google.com/vertex-ai/pricing>
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning with applications in R (2nd ed.). Springer.
- Jaramillo, C. (2022). *Influencia del empleo estudiantil en el rendimiento académico de las instituciones educativas para personas con escolaridad inconclusa (PCEI) de la ciudad de Otavalo (Master's thesis, Otavalo)*.  
<https://repositorio.uotavalo.edu.ec/bitstream/52000/763/1/PP-EDU2-2022-060.pdf>
- Mejía, C. (2023). *M-Learning y el aprendizaje autónomo de matemática de los estudiantes de básica superior (Master's thesis, Pontificia Universidad Católica del Ecuador)*.

<https://repositorio.puce.edu.ec/server/api/core/bitstreams/7c37a00d-e387-4d05-b142-2ba2bf6d3a61/content>

Microsoft. (2024). Power BI Pricing. Recuperado el 10 de marzo de 2025, de <https://powerbi.microsoft.com/pricing/>

Monteros, E. (2024). *vivencial de aprendizaje para el desarrollo de la lectura comprensiva en estudiantes con escolaridad inconclusa (Master's thesis, Quito: Universidad Tecnológica Indoamérica)*.  
<http://repositorio.uti.edu.ec/bitstream/123456789/6497/1/MONTEROS%20VACA%20EDISON%20LEONARDO.pdf>

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th ed.). Wiley.

ProjectPro. (2024). Google Cloud Dataflow vs. Apache Spark. , <https://www.projectpro.io/compare/google-cloud-dataflow-vs-apache-spark>

Reinoso, H. (2024). *Análisis de funciones y responsabilidades de un gestor educativo en centros de escolaridad inconclusa: Analysis of functions and responsibilities of an educational manager in unfinisheds*. . <https://latam.redilat.org/index.php/lt/arti>

Tumbaco, M. (2022). *Impacto del programa educativo campaña todos ABC y aprendizaje en estudiantes con escolaridad inconclusa, en una unidad educativa, Santa Elena*.  
[https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/93770/Tumbaco\\_MM\\_C-SD.pdf](https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/93770/Tumbaco_MM_C-SD.pdf)


## ANEXOS

### ANEXO 1

#### FORMATO DE ENCUESTA

# Encuesta sobre Escolaridad Inconclusa en Estudiantes

Esta encuesta busca recopilar información para analizar factores que influyen en la deserción escolar y desarrollar un modelo predictivo basado en datos.

 No compartido

**\* Indica que la pregunta es obligatoria**

¿Cuál es su edad? \*

Tu respuesta

¿Cuál es su género? \*

- Masculino
- Femenino
- Prefiero no decirlo

¿En qué tipo de institución estudia/estudió? \*

- Pública
- Privada
- Fiscal-mixta
- Fiscomisional
- Municipal

¿En qué nivel educativo se encuentra actualmente o dejó sus estudios? \*

- Primaria
  - Secundaria
  - Bachillerato
  - Educación Superior
  - No aplica
- 

Cuál es el nivel educativo más alto alcanzado por sus padres o tutores? \*

- Primaria
  - Secundaria
  - Bachillerato
  - Universitario
  - Posgrado
  - No lo sé
- 

Cuál es el ingreso mensual aproximado de su hogar? \*

- Menos de \$200
- \$200-\$400
- \$400-\$600
- \$600-\$1000
- Más de \$1000
- Prefiero no decir

¿Trabaja actualmente o ha trabajado mientras estudiaba? \*

- Sí
  - No
- 

Si trabaja, ¿cuántas horas semanales dedica al trabajo? \*

- Menos de 10
  - 10-20
  - 20-30
  - Más de 30
  - No aplica
- 

¿Recibe algún tipo de apoyo económico o beca para estudiar? \*

- Sí
  - No
- 

¿Ha considerado dejar sus estudios en algún momento? \*

- Sí
  - No
- 

Si dejó los estudios, ¿cuál fue el motivo principal? \*

- Problemas económicos
- Falta de interés
- Bajo rendimiento académico
- Problemas familiares
- Razones de salud
- Necesidad de trabajar
- Otra

¿Cuántas veces ha repetido un curso o grado escolar? \*

- Ninguna
- 1 vez
- 2 veces
- Más de 2 veces

¿Se siente motivado para continuar con sus estudios? \*  
(Escala de 1 a 5, donde 1 es nada motivado y 5 es muy motivado)

- 1
- 2
- 3
- 4
- 5

¿Cuenta con acceso a internet y dispositivos electrónicos para estudiar? \*

- Sí
- No

¿Ha tomado clases virtuales o a distancia en los últimos dos años? \*

- Sí
- No

Enviar

Borrar formulario

## ANEXO 2

### DATOS ABIERTOS DEL MINISTERIO DE EDUCACION DEL ECUADOR

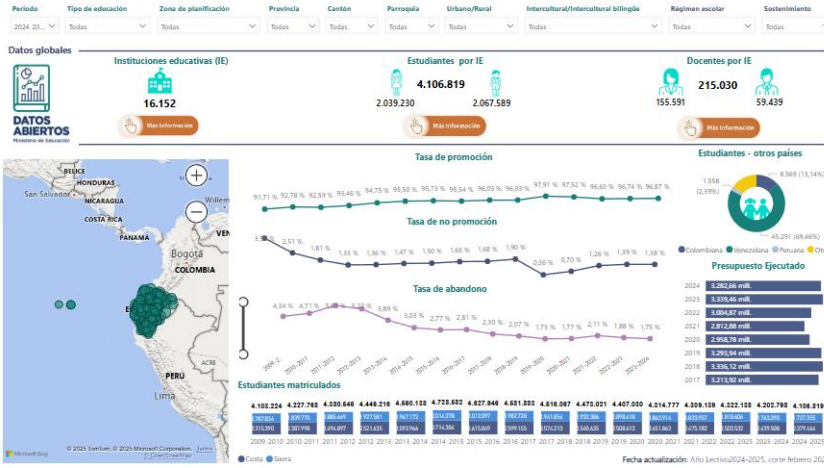
- Visualizador de datos
- Base de datos
- Información educativa
- Fichas metodológicas
- Publicaciones estadísticas
- Investigación educativa

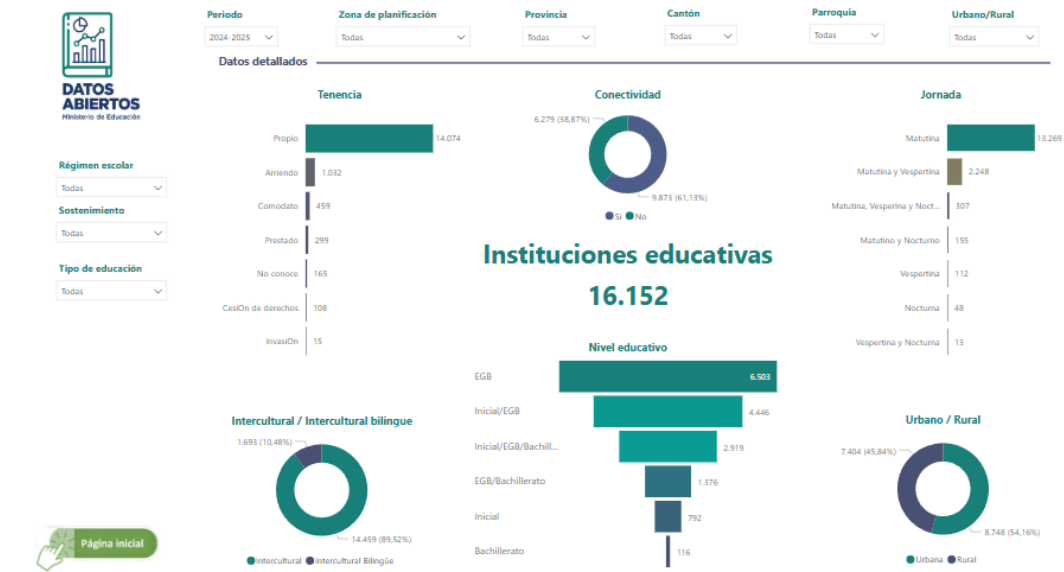
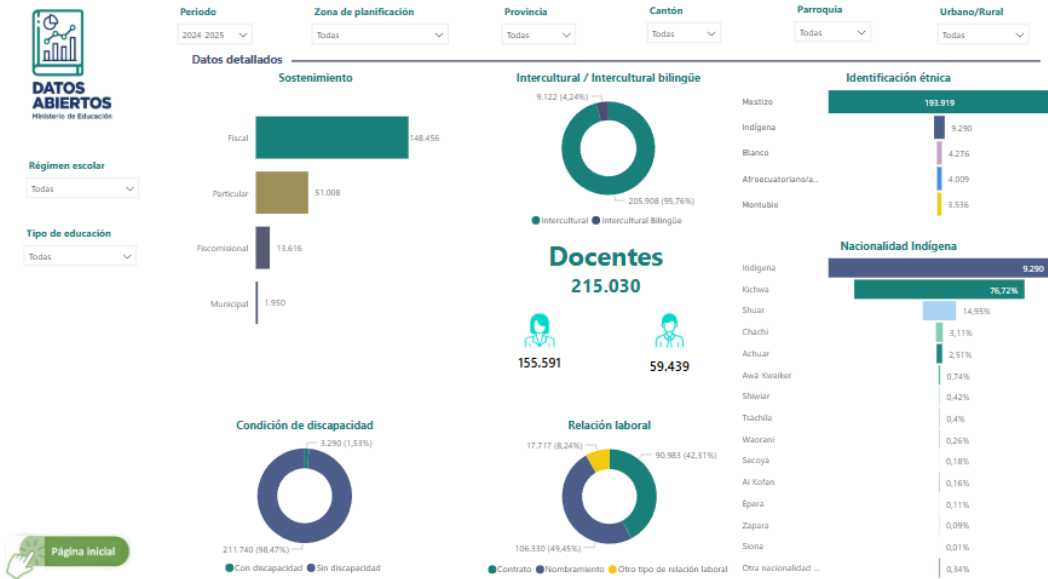


**DATOS ABIERTOS**  
Ministerio de Educación

Presentación de CIFRAS HISTÓRICAS EDUCATIVAS de fácil manejo y de acceso abierto.







### ANEXO 3

#### VALIDACIÓN DE EXPERTOS



## UNIVERSIDAD TECNOLÓGICA ISRAEL

### ESCUELA DE POSGRADOS "ESPOG"

#### MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

#### INSTRUMENTO PARA VALIDACIÓN DE LA PROPUESTA

Estimado colega:

Se solicita su valiosa cooperación para evaluar la calidad del siguiente contenido digital "Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación". Sus criterios son de suma importancia para la realización de este trabajo, por lo que se le pide que brinde su cooperación contestando las preguntas que se realizan a continuación.

#### Datos informativos

|  |
|--|
| Validado por: Ing. Tonysé de la Rosa Martín, Ms.C  |
| Título obtenido: Máster en Bioinformática y Biología Computacional   |
| C.I.: 1756806814   |
| E-mail: severus.trm@gmail.com  |
| Institución de Trabajo: Universidad Iberoamericana del Ecuador(Pregrado y Postgrados),<br>Universidad Tecnológica Israel(Postgrados) |
| Cargo: Docente-Investigador  |
| Años de experiencia en el área: 20   |

**Instructivo:**

- Responda cada criterio con la máxima sinceridad del caso.
- Revisar, observar y analizar la propuesta de la plataforma virtual, blog o sitio web.
- Coloque una X en cada indicador, tomando en cuenta que Muy adecuado equivale a 5, Bastante Adecuado equivale a 4, Adecuado equivale a 3, Poco Adecuado equivale a 2 e Inadecuado equivale a 1.

**Tema: “Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación”**

| Indicadores                | Muy adecuado | Bastante Adecuado | Adecuado | Poco adecuado | Inadecuado |
|----------------------------|--------------|-------------------|----------|---------------|------------|
| Pertinencia                | x            |                   |          |               |            |
| Aplicabilidad              | x            |                   |          |               |            |
| Factibilidad               | x            |                   |          |               |            |
| Novedad                    |              | x                 |          |               |            |
| Fundamentación pedagógica  | x            |                   |          |               |            |
| Fundamentación tecnológica | x            |                   |          |               |            |
| Indicaciones para su uso   | x            |                   |          |               |            |
| <b>TOTAL</b>               | <b>6</b>     | <b>1</b>          |          |               |            |

**Observaciones:** Este tipo de soluciones son muy importantes en el ámbito de la Educación actual, dado que permiten tomar decisiones efectivas sobre un tema muy importante como la no conclusión de estudios. Basadas en técnicas de Machine Learning, constituye una novedosa solución a problemas del ministerio de educación y extensible a otras áreas de la vida.

**Recomendaciones:** Se recomienda Publicar los resultados en Papper Científico para revistas indexadas de Clase Scopus en el área de las Ciencias de la Computación y afines.

**Lugar, fecha de validación:** Quito, 8 de marzo de 2025



Ing. Tonyse de la Rosa Martín, Ms.C

**ANEXO 4**  
**VALIDACIÓN DE EXPERTOS**



**UNIVERSIDAD TECNOLÓGICA ISRAEL**  
**ESCUELA DE POSGRADOS “ESPOG”**

**MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS**

**INSTRUMENTO PARA VALIDACIÓN DE LA PROPUESTA**

Estimado colega:

Se solicita su valiosa cooperación para evaluar la calidad del siguiente contenido digital “Ejecución de un Modelo de Machine Learning para el Análisis de la Escolaridad Inconclusa en Estudiantes del Ministerio de Educación”. Sus criterios son de suma importancia para la realización de este trabajo, por lo que se le pide que brinde su cooperación contestando las preguntas que se realizan a continuación.

**Datos informativos**

|  |  |
|--|--|
| <b>Validado por:</b>                   | Pablo Marcel Recalde V   |
| <b>Título obtenido:</b>                | Magíster MSc. En Gestión de las Comunicaciones y Tecnologías de la Información |
| <b>C.I.:</b>                           | 1711685055   |
| <b>E-mail:</b>                         | precalde@uisrael.edu.ec  |
| <b>Institución de Trabajo:</b>         | Universidad Israel   |
| <b>Cargo:</b>                          | Coordinador de Carreras  |
| <b>Años de experiencia en el área:</b> | Veinte (20)  |

**Instructivo:**

- Responda cada criterio con la máxima sinceridad del caso.
- Revisar, observar y analizar la propuesta de la plataforma virtual, blog o sitio web.
- Coloque una X en cada indicador, tomando en cuenta que Muy adecuado equivale a 5, Bastante Adecuado equivale a 4, Adecuado equivale a 3, Poco Adecuado equivale a 2 e Inadecuado equivale a 1.

Tema: “ ”

| Indicadores                | Muy adecuado | Bastante Adecuado | Adecuado | Poco adecuado | Inadecuado |
|----------------------------|--------------|-------------------|----------|---------------|------------|
| Pertinencia                | x            |                   |          |               |            |
| Aplicabilidad              | x            |                   |          |               |            |
| Factibilidad               |              | x                 |          |               |            |
| Novedad                    | x            |                   |          |               |            |
| Fundamentación pedagógica  |              | x                 |          |               |            |
| Fundamentación tecnológica | x            |                   |          |               |            |
| Indicaciones para su uso   | x            |                   |          |               |            |
| <b>TOTAL</b>               | <b>20</b>    | <b>8</b>          |          |               |            |

**Observaciones:** Sería más adecuado incluir fundamentación estadística y de procesos para analizar la data que se presenta, no obstante, la pertinencia es muy adecuada.

**Recomendaciones:** Incluir mayor sustento matemático del porque elegir los tipos de algoritmos utilizados.

**Lugar, fecha de validación:** Quito, 11 de marzo de 2025



Autenticado digitalmente por:  
PABLO MARCEL  
RECALDE VARELA

MSc. Ing. Pablo Recalde