

## “LUMINA: Un Agente de IA para la Clasificación de Perfiles Docentes de Retroalimentación”

Sheny San Martín Quiroga

CRUCH Bio Bío Ñuble

[shsanmartin@doctoradoia.cl](mailto:shsanmartin@doctoradoia.cl)

Inteligencia Artificial

Noviembre 2025

### Resumen (Abstract)

Se presenta LUMINA, un agente inteligente que utiliza Regresión Logística Multinomial para clasificar a docentes en tres niveles de apropiación de retroalimentación efectiva (Bajo, Medio, Alto) a partir de 10 indicadores normalizados. La elección del modelo se justifica por su capacidad predictiva y su alta interpretabilidad, lo que facilita la toma de decisiones pedagógicas. El modelo fue entrenado y validado con una muestra de 250 docentes, obteniendo un *Accuracy* global de 0.91, *F1-score* macro de 0.95 y *AUC-ROC* promedio de 0.99. La aplicación web interactiva permite a cualquier docente autoevaluar su práctica en menos de 1 minuto y recibir una predicción junto con recomendaciones pedagógicas. Los resultados demuestran que el uso y frecuencia de herramientas digitales y el acceso a formación específica son los predictores más significativos, orientando las futuras intervenciones de desarrollo profesional. El sistema se propone como una herramienta de bajo costo y alta interpretabilidad para apoyar procesos de desarrollo profesional docente continuo.

*Palabras claves:* Agente Inteligente, Clasificación Multiclasa, Educación, Machine Learning, Retroalimentación docente.

## 1 Introducción

### 1.1 Descripción del Problema

La retroalimentación ha sido consistentemente identificada en la literatura especializada como un componente esencial en el proceso educativo, reconocido por su capacidad para promover la reflexión crítica y mejorar el aprendizaje (Amaranti Pesce & Contreras Pérez, 2025).

Conceptualmente, se define como la información entregada sobre el desempeño o comprensión de una tarea, cuyo fin último es cerrar la brecha entre el conocimiento actual del estudiante y la meta deseada (Lira et al., 2021). Diversos estudios respaldan que su efectividad impacta significativamente el rendimiento académico, superando incluso los factores socioeconómicos, reportando un tamaño del efecto de 0.73 según el metaanálisis de Hattie y Timperley (2007).

No obstante, la evidencia empírica advierte que la mera presencia de retroalimentación no garantiza efectos positivos. Su impacto puede ser nulo o incluso contraproducente si el enfoque es incorrecto, por ejemplo, cuando se centra en la persona del estudiante (“eres bueno”) en lugar de en el proceso de la tarea o la autorregulación (Hattie & Timperley, 2007).

Esta problemática se agudiza en el contexto nacional chileno. A pesar de que las políticas educativas, impulsada por el Decreto 67 (2018) del Ministerio de Educación, exigen y promueven la implementación de la retroalimentación formativa y continua, la evidencia en la práctica en el aula revela una brecha crítica.

En este escenario, el rol del docente se vuelve fundamental. Aunque la literatura ha explorado extensamente qué constituye una “buena retroalimentación”, existe un menor desarrollo respecto a cómo los docentes se apropián de este constructo. La apropiación docente debe entenderse no sólo como la adquisición del conocimiento teórico, sino como la integración de este saber en la acción pedagógica, considerando cómo el contexto laboral modulas dichas prácticas.

Identificar estos niveles de apropiación de forma manual (a través de la observación o encuestas subjetivas) resulta costoso, inescalable y propenso a sesgos. En respuesta a esta necesidad, la presente investigación aborda esta brecha específica desde una perspectiva tecnológica. Situada en el contexto de la educación media chilena (de 7º básico a 4º medio), la propuesta central consiste en el diseño y validación de un agente de inteligencia artificial. A diferencia de estudios puramente descriptivos, este trabajo busca instrumentalizar el “Índice de Apropiación de la Retroalimentación Docente” (IARD) para entrenar un modelo predictivo. El objetivo es que dicho agente sea capaz de clasificar automáticamente el nivel de apropiación docente, facilitando una caracterización objetiva y escalable. Considerando estos antecedentes, surge la siguiente pregunta de investigación que guía este estudio: ¿De qué manera un modelo de Inteligencia Artificial puede caracterizar y predecir los distintos niveles de apropiación de la retroalimentación en docentes de Enseñanza Media?

## 1.2 Estado del Arte

La literatura contemporánea ha avanzado significativamente en la comprensión de la retroalimentación docente, destacándola como un elemento esencial para promover la reflexión crítica y mejorar los aprendizajes. Estudios como los de Hattie y Timperley (2007) muestran que la retroalimentación posee un alto potencial para influir positivamente en el rendimiento académico. Sin embargo, estos mismos autores advierten que su impacto puede ser nulo o incluso negativo cuando se centra en la persona del estudiante en lugar de focalizarse en la tarea, los procesos cognitivos o la autorregulación. Este enfoque abre un problema central: aunque la investigación se ha concentrado en analizar la calidad del mensaje entregado por el docente (*input*), sigue existiendo un vacío respecto a cómo medir la apropiación y el efecto real de dicho feedback en el estudiante (*output*).

En los últimos años, la percepción que tienen los estudiantes sobre la enseñanza y, particularmente, sobre la retroalimentación, ha adquirido un rol fundamental como indicador de calidad educativa. Diversos estudios han demostrado que estas valoraciones no son un mecanismo meramente administrativo, sino un predictor relevante del desempeño académico. Hernández Medina, Ramírez Torres y Pinilla Rodríguez (2024) evidencian que las evaluaciones positivas del profesorado por parte

de los estudiantes se correlacionan directamente con el rendimiento académico, confirmando la validez de estas percepciones en modelos de estimación del desempeño. De manera complementaria, Ramos Silva y Carvalho (2021) muestran que la percepción de prácticas eficaces de feedback docente se asocia con mejores resultados académicos. Asimismo, investigaciones como las de Ríos Muñoz y Herrera Araya (2023) indican que los estudiantes valoran positivamente la retroalimentación sistemática, oportuna y constructiva, reforzando la importancia de considerar su perspectiva para evaluar la efectividad del proceso educativo.

En paralelo, el desarrollo de la Inteligencia Artificial (IA) ha transformado profundamente las prácticas pedagógicas, con dos líneas sobresalientes en la literatura. Por un lado, los Sistemas Tutores Inteligentes y el aprendizaje adaptativo permiten personalizar trayectorias educativas mediante algoritmos que ajustan el contenido, el ritmo o los ejercicios según el desempeño individual del estudiante, proporcionando retroalimentación automática e inmediata (Hwang et al., 2020). Por otro lado, el campo de Learning Analytics ha demostrado el potencial de la IA, especialmente del aprendizaje automático, para analizar grandes volúmenes de datos educativos y generar predicciones sobre riesgo académico, abandono, patrones de aprendizaje o comportamientos estudiantiles (Baker & Siemens, 2014). Aunque estos avances han impulsado la toma de decisiones informadas en educación, la mayoría se ha enfocado en resultados estudiantiles, dejando menos desarrollados los modelos centrados en el desempeño docente y, particularmente, en su práctica de retroalimentación.

En el contexto chileno, las experiencias con IA en educación han emergido principalmente a través de proyectos piloto impulsados por el Centro de Innovación del Ministerio de Educación y por instituciones de educación superior. También se observa una creciente adopción de plataformas internacionales basadas en IA, adaptadas al currículo nacional. No obstante, persiste una ausencia de modelos que permitan evaluar y clasificar perfiles docentes asociados específicamente a la retroalimentación, utilizando datos provenientes de la percepción estudiantil y técnicas de aprendizaje automático.

En síntesis, aunque la literatura reconoce la retroalimentación como una práctica determinante para el aprendizaje y ha avanzado en sistematizar la percepción estudiantil como indicador de calidad docente, aún existe un vacío en cuanto a herramientas que permitan modelar y clasificar perfiles de retroalimentación. Este vacío justifica la propuesta de LUMINA, un modelo de IA destinado a identificar y caracterizar estos perfiles a partir de indicadores estandarizados, ofreciendo así una aproximación innovadora para comprender la efectividad de la retroalimentación en el aula y su potencial influencia en el aprendizaje.

### **1.3 Objetivos de la Investigación**

#### **1.3.1 Objetivo general**

Diseñar e implementar un agente inteligente basado en Machine Learning capaz de predecir el nivel de apropiación de retroalimentación efectiva en docentes y entregar recomendaciones pedagógicas personalizadas.

#### **1.3.2 Objetivos específicos**

1. Analizar y preprocessar un conjunto de datos.

2. Construir el Índice de Apropiación de la Retroalimentación Docente (IARD), que permitirá medir el nivel de conocimiento, práctica y uso de tecnología en torno a la retroalimentación.
3. Implementar y evaluar un modelo de Machine Learning.
4. Identificar los predictores más significativos para la apropiación de retroalimentación.
5. Diseñar un agente capaz de predecir el nivel de apropiación, para ofrecer apoyo pedagógico.

## **2 Metodología**

### **2.1 Diseño y Contexto del Estudios**

La presente investigación se desarrolló bajo un enfoque cuali-cuantitativo, de tipo no experimental y transversal, orientado a identificar el IARD y determinar el perfil del profesorado que implementa prácticas efectivas de retroalimentación en el aula.

Los datos fueron recopilados de forma automática mediante un formulario en línea durante los meses de agosto a noviembre y posteriormente fueron exportados a una hoja de cálculo para su análisis. El cuestionario fue distribuido electrónicamente a través de un enlace y código QR, enviados por correo electrónico a los Jefes de UTP, quienes facilitaron su aplicación entre los docentes en ejercicio de sus establecimientos. La participación fue voluntaria y anónima, garantizando la confidencialidad de la información recolectada. No se realizó ninguna intervención ni manipulación experimental.

### **2.2 Participantes**

La muestra estuvo compuesta por 250 docentes de educación básica y media, específicamente desde séptimo año de enseñanza básica hasta cuarto año medio, pertenecientes a distintos establecimientos educacionales chilenos. La selección se realizó mediante un muestreo intencionado, invitando a los docentes a participar de manera voluntaria no remunerada y se basó en el consentimiento informado. Los docentes accedieron a través de enlace seguro en línea y la completaron individualmente, utilizando como criterio de exclusión que los participantes estuvieran ejerciendo la docencia en los niveles antes mencionados. El promedio de edad fue de 41.78 años, y las edades de los participantes oscilaron entre 23 y 72 años y con un promedio de carga horaria de 34 horas semanales, donde el 58,4% pertenece a colegios municipales y el 41,6% a otras dependencias.

### **2.3 Instrumento**

El instrumento de recolección de datos consistió en un cuestionario estructurado elaborado en un formulario de Google, diseñado a partir de una revisión exhaustiva de literatura sobre retroalimentación docente en el aula y sobre factores asociados a su apropiación pedagógica. La versión inicial estuvo compuesta por 38 ítems, los cuales fueron sometidos a un proceso de validación de contenido mediante juicio de expertos. Estos evaluaron cada ítem en términos de suficiencia, claridad, coherencia y relevancia respecto de los objetivos de la investigación. Como resultado de sus observaciones, se realizaron ajustes y depuración, obteniéndose una versión final de 36 preguntas, organizadas en cinco dimensiones:

- Dimensión 1: Datos sociodemográficos (14)  
Caracteriza el perfil profesional y laboral del docente.
- Dimensión 2: Conocimiento sobre retroalimentación (3)  
Evalúa el nivel de formación, conceptualización y valoración de la retroalimentación.
- Dimensión 3: Prácticas de retroalimentación (7)  
Indaga en momentos, técnicas, enfoques y tiempos dedicados a la retroalimentación en el aula.
- Dimensión 4: Uso de tecnologías (4)  
Explora frecuencia, tipo de herramientas tecnológicas utilizadas y percepciones sobre su utilidad y dificultades.
- Dimensión 5: Percepción de la institucionalidad (7)  
Considera aspectos como cultura escolar, apoyo institucional, integración formal de la retroalimentación y capacitaciones.

El cuestionario combinó ítems de tipo cerrado, principalmente escalas ordinales tipo Likert, y preguntas de selección múltiple, con preguntas abiertas, lo que permitió captar tanto la intensidad y frecuencia de las prácticas docentes como aspectos cualitativos sobre experiencias y desafíos. El conjunto de datos resultante corresponde a un tipo corte transversal, compuesto por 36 variables que abarcan el perfil sociodemográfico, el conocimiento, las prácticas pedagógicas, el uso de tecnologías y la percepción institucional relacionada con la retroalimentación. El dataset contiene predominantemente variables categóricas (31 variables, 86.1 % del total), acompañadas de un número reducido de variables numéricas (3 variables, 8,3 %) y respuestas cualitativas abiertas (2 variables, 5.6 %) (Tabla 1).

Para la construcción del Índice de Apropiación de la Retroalimentación Docente (IARD), se seleccionaron las variables pertenecientes a las dimensiones de Conocimiento, Prácticas de retroalimentación y Uso de tecnologías, ya que estas representan los componentes que la literatura vincula más directamente con la apropiación pedagógica del concepto. Previo al cálculo del índice, dichas variables fueron normalizadas mediante MinMaxScaler, con el objetivo de reflejar consistentemente la calidad y alineación de las prácticas reportadas con los principios de retroalimentación formativa.

## 2.4 Procedimiento

Para obtener el Índice de Apropiación de la Retroalimentación Docente (IARD), se desarrolló un procedimiento de cuatro etapas: selección de variables, preprocesamiento de datos, normalización de puntuaciones y cálculo del índice. Cada fase se detalla a continuación.

### 2.4.1 Selección de variables relevantes

Con base en la literatura sobre retroalimentación formativa, se seleccionaron variables pertenecientes a tres dimensiones del cuestionario:

- Conocimiento sobre retroalimentación
- Prácticas de retroalimentación en el aula
- Uso de tecnologías para retroalimentar

Estas dimensiones representan los componentes que influyen directamente en el nivel de apropiación del docente, ya que integran aspectos cognitivos, procedimentales y tecnológicos asociados a la retroalimentación. Uso de tecnologías para la retroalimentación Conocimiento sobre la retroalimentación.

Las variables sociodemográficas y las relacionadas a la institucionalidad fueron excluidas del índice por no constituir conductas o conocimientos atribuibles directamente al docente.

#### **2.4.2 Preprocesamiento del conjunto de datos**

El proceso de preprocesamiento consistió en codificar las variables y revisar los valores faltantes.

Las variables con respuestas en escala Likert se codificaron numéricamente en un rango ascendente (por ejemplo: Nunca = 0, A veces = 1, Frecuentemente = 2, Siempre = 3).

Esta codificación permitió cuantificar el grado de alineación de cada docente con las prácticas de retroalimentación formativa.

Las preguntas de selección múltiple se contabilizaron y se transformaron a un valor numérico.

Las preguntas abiertas fueron excluidas del cálculo del índice.

#### **2.4.3 Normalización de las puntuaciones**

Con el objetivo de que todas las variables tuvieran el mismo peso en el cálculo del índice y que sus escalas fueran comparables, se aplicó una normalización MinMaxScaler:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Esta transformación generó un conjunto de puntuaciones entre 0 y 1, donde:

- **0** representa el nivel más bajo de apropiación.
- **1** representa el nivel más alto.

Este paso es crucial, ya que las variables del cuestionario tenían escalas y formatos diferentes, y una comparación directa no habría sido válida sin una estandarización previa.

#### **2.4.4 Índice de Apropiación de la Retroalimentación Docente (IARD)**

Una vez normalizadas las variables se calculó el índice mediante un promedio ponderado simple:

$$\text{IARD}_{\text{global}} = \frac{1}{n} \sum_{i=1}^n X_{\text{norm},i}$$

Donde:

$n$  = número total de variables seleccionadas de las tres dimensiones.

$X_{\text{norm}, i}$  = valor normalizado de cada variable.

Este método otorga el mismo peso a cada indicador, asumiendo que todas las dimensiones contribuyen por igual a la apropiación.

#### **2.4.5 Clasificación de niveles de apropiación**

Para la interpretación del índice se establecieron tres niveles, definidos a partir de los cuartiles del conjunto de datos:

- **Bajo nivel** de apropiación: 0,00 – 0,33
- **Medio nivel** de apropiación: 0,34 – 0,66
- **Alto nivel** de apropiación: 0,67 – 1,00

Esta clasificación permitió identificar patrones y agrupar a los docentes según su grado de alineamiento con las prácticas de retroalimentación formativa.

A partir de las variables normalizadas seleccionadas y del nivel de apropiación obtenido mediante el IARD, se construyó un nuevo dataset específicamente diseñado para el entrenamiento del modelo predictivo. Este conjunto de datos incluye:

- **Variables independientes (predictoras):** se utilizaron todas las variables pertenecientes a las dimensiones de Conocimiento, Prácticas de retroalimentación y Uso de tecnologías y que fueron normalizadas en un rango de 0 a 1.
- **Variable objetivo:** *nivel\_apropiación\_retroalimentación*, una variable categórica multiclasé derivada del índice global, con la siguiente distribución en la muestra:
  - **Bajo:** 28 %
  - **Medio:** 52 %
  - **Alto:** 20 %

Este dataset final constituye la base para el proceso de entrenamiento y validación del modelo de clasificación, permitiendo evaluar la capacidad de la inteligencia artificial para predecir el nivel de apropiación docente en función de sus prácticas, conocimientos y uso de tecnologías.

### **2.5 Algoritmo de Machine Learning**

Para la predicción del nivel de apropiación de la retroalimentación docente se empleó el algoritmo Multinomial Logistic Regression (MLR). Este modelo, implementado mediante la clase LogisticRegression de *scikit-learn* con *multi\_class='multinomial'* y el solver *lbfgs*, fue seleccionado tras evaluar alternativas como Random Forest, Support Vector Machine, entre otros, priorizando criterios de interpretabilidad, robustez estadística y alineación conceptual con los objetivos del estudio.

#### **2.5.1 Justificación del Algoritmo**

##### **a) Alta interpretabilidad**

A diferencia de modelos más complejos como Random Forest o SVM, la Regresión Logística Multinomial permite interpretar directamente los coeficientes asociados a cada predictor, lo que hace posible:

- Identificar la dirección del efecto (incremento o disminución del odds).
- Estimar la magnitud relativa de la influencia de cada variable.
- Explicar de forma transparente por qué el modelo predice un nivel de apropiación específico.

En un contexto educativo, especialmente en investigación sobre prácticas docentes, la interpretabilidad es un requisito central, ya que los resultados deben poder traducirse en acciones formativas y decisiones pedagógicas fundamentadas.

**b) Adecuación al tipo de variable objetivo**

La variable dependiente es categórica multiclase (Bajo, Medio, Alto).

La MLR está diseñada para este tipo de problemas y modela explícitamente las probabilidades de pertenecer a cada categoría, lo que favorece:

- una interpretación probabilística,
- un análisis más fino de los límites entre categorías,
- y la posibilidad de identificar incertidumbre en la clasificación.

**c) Estabilidad y bajo riesgo de sobreajuste**

Dado que el dataset no es extremadamente grande, un modelo paramétrico y regularizado como la MLR evita el sobreajuste que puede producirse con métodos altamente complejos.

**2.5.2 Proceso de Entrenamiento y Optimización**

Todo el procedimiento fue implementado en el archivo *train.py* mediante la librería *scikit-learn*, siguiendo las siguientes etapas:

**a) Selección de Características**

Para asegurar que el modelo utilizará únicamente variables relevantes y evitar redundancias, se aplicó un proceso de selección de características. El conjunto inicial estaba compuesto por 17 variables normalizadas (0–1), seleccionadas teóricamente desde las dimensiones de Conocimiento, Prácticas de retroalimentación y Uso de tecnologías.

La identificación de las variables más influyentes se realizó mediante Permutation Importance (*scikit-learn*), técnica que evalúa cuánto disminuye el desempeño del modelo al permutar aleatoriamente cada predictor. Una caída mayor en la métrica (accuracy) indica mayor importancia. Este método fue seleccionado por su interpretabilidad y porque evalúa directamente el aporte de cada variable sobre el modelo entrenado.

Los análisis mostraron que 10 de las 17 variables generaban una disminución significativa en el desempeño al ser permutadas, por lo que fueron retenidas. Las variables con impacto mínimo fueron descartadas, conformando así el dataset final usado para entrenar el modelo. Esto permitió obtener un modelo más simple, eficiente e interpretativo sin sacrificar rendimiento.

**b) División del dataset**

Se realizó una división estratificada, destinando:

- **70 %** para entrenamiento
- **30 %** para prueba

El uso de *StratifiedShuffleSplit* garantizó que la distribución original de clases (Bajo 28%, Medio 52%, Alto 20%) se mantuviera en ambos subconjuntos, evitando sesgos y variaciones artificiales en la proporción de niveles.

**c) Balanceo de clases durante el entrenamiento**

Debido a que la clase Medio es mayoritaria en la muestra, se utilizó el parámetro:

*class\_weight= 'balanced'*

Este ajuste asigna pesos inversamente proporcionales a la frecuencia de cada clase en la función de pérdida, con dos efectos importantes:

- Evita que el modelo aprenda a privilegiar la clase mayoritaria.

- Mejora la sensibilidad a las clases Bajo y Alto, que son menos representadas pero cruciales para el análisis educativo.

**d) Optimización del hiperparámetro de regularización**

El principal hiperparámetro de la MLR es:

$$C = \frac{1}{\lambda}$$

donde  $C$  controla la fuerza de la regularización (valores mayores → menor regularización).

Se aplicó una búsqueda en cuadrícula (Grid Search) con validación cruzada estratificada (StratifiedKFold), explorando diferentes valores posibles para  $C$ .

El valor óptimo obtenido fue:

$C=10$

Este punto logró el mejor equilibrio entre:

- ajuste del modelo,
- estabilidad,
- generalización.

**e) Entrenamiento del modelo final**

El modelo final se entrenó con:

- `multi_class='multinomial'`
- `solver='lbfgs'`
- `class_weight='balanced'`
- $C=10$

Tras el entrenamiento, los coeficientes del modelo permitieron identificar las variables predictoras con mayor influencia, lo cual contribuye al análisis interpretativo del comportamiento docente.

## 2.6 Diseño del Agente

El agente desarrollado corresponde a un sistema inteligente de apoyo a la toma de decisiones, diseñado para predecir el nivel de apropiación de la retroalimentación docente a partir de un conjunto de respuestas autoinformadas. Su arquitectura combina componentes de Machine Learning, procesamiento de datos y presentación interactiva, organizados en un flujo secuencial y coherente.

### 2.6.1 Arquitectura del Agente

El agente está compuesto por los siguientes módulos principales:

**a) Módulo de Interfaz (Front-End Interactivo)**

Implementado en Streamlit, permite que el usuario responda una serie de afirmaciones mediante escalas Likert. Esta interfaz captura la información en tiempo real y la transforma en variables normalizadas compatibles con el modelo.

**b) Módulo de Preparación de Datos**

- Construye un vector de características basado en las respuestas del usuario.
- Ordena las variables según el listado original de características utilizadas para el entrenamiento.

- Completa los valores no entregados con ceros y aplica la transformación mediante el escalador MinMaxScaler previamente entrenado.

c) **Módulo de Inferencia con el Modelo Predictivo**

Incluye el modelo de Regresión Logística Multinomial, el cual:

- Recibe el vector normalizado del usuario.
- Genera una predicción de nivel de apropiación (Bajo, Medio o Alto).
- Calcula las probabilidades asociadas a cada clase.
- Reconstruye la clase nominal utilizando un *LabelEncoder*.

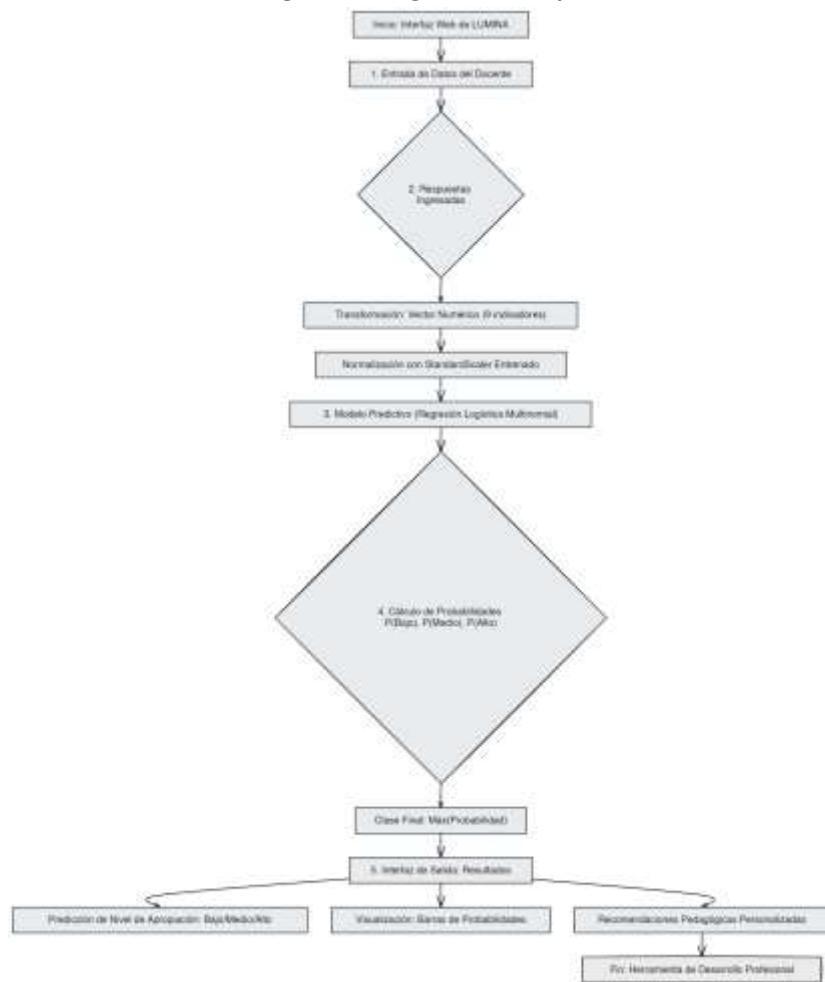
d) **Módulo de Interpretación y Retroalimentación**

El agente no solo entrega la clase predicha, sino también:

- Probabilidad asociada a cada nivel.
- Un mensaje adaptativo según el resultado.
- Recomendaciones personalizadas.
- Visualización gráfica mediante un gráfico de barras (matplotlib).

e) **Flujo de Trabajo del Agente**

Figura 1: Diagrama de flujo



Este flujo permite una interacción natural, rápida y comprensible para cualquier docente o directivo.

## 2.7 Implementación Técnica del Agente

La implementación del agente se realizó en Python, integrando librerías especializadas para Machine Learning, manejo de datos y desarrollo de interfaces interactivas. El sistema combina un modelo previamente entrenado con una interfaz accesible que permite realizar predicciones en tiempo real.

### 2.7.1 Tecnologías y Herramientas Utilizadas

Tabla 1: Herramientas y sus usos en LUMINA

Herramientas	Uso
<b>Streamlit</b>	Utilizado como framework para la interfaz web del agente. Permite: <ul style="list-style-type: none"><li>• Formular preguntas.</li><li>• Capturar respuestas.</li><li>• Mostrar resultados y visualizaciones.</li><li>• Gestionar estado entre interacciones.</li></ul>
<b>scikit-learn</b>	Usado para: <ul style="list-style-type: none"><li>• Cargar el modelo entrenado (LogisticRegression).</li><li>• Aplicar el escalador (MinMaxScaler).</li><li>• Decodificar clases (LabelEncoder).</li></ul>
<b>pickle</b>	Permite cargar el modelo entrenado en formato serializado, incluyendo: <ul style="list-style-type: none"><li>• Algoritmo predictivo.</li><li>• Escalador.</li><li>• Codificador de etiquetas.</li><li>• Lista de características requeridas.</li></ul>
<b>NumPy y pandas</b>	Utilizados para la construcción y manipulación del vector de entrada.
<b>matplotlib</b>	Empleado para generar la visualización gráfica de las probabilidades de clasificación.

### 3 Resultados

#### 3.1 Evaluación del Modelo

El modelo de Regresión Logística Multinomial mostró un desempeño sólido y estable al ser evaluado en el conjunto de prueba (30 % de los datos). Las métricas globales se presentan en la siguiente tabla:

Tabla 2: Métricas generales

Métrica	Valor
Accuracy (Global)	0.9067
Precision (macro)	0.9051
Recall (macro)	0.9083
F1-score (macro)	0.9495
AUC-ROC (OvR Promedio)	0.9941

El F1-score macro (0.95) es particularmente relevante, ya que entrega una medida equilibrada del desempeño en cada clase, incluso cuando existe un leve desbalance. Un valor superior a 0.80 refleja una excelente capacidad de clasificación. Por su parte, el valor de AUC-ROC (0.99) confirma que el modelo logra una separación clara entre las clases Bajo, Medio y Alto, evidenciando su capacidad discriminativa.

#### 3.2 Reporte de Clasificación por Clase (Conjunto de Prueba)

Tabla 3: Métricas por clase

Clase	Precisión	Recall	F1-score	Soporte
Bajo	1.00	0.88	0.93	24
Medio	0.89	0.92	0.90	36
Alto	0.82	0.93	0.88	15

El desempeño es consistente en las tres clases. La clase Medio presenta los mejores indicadores, lo cual es esperable dada su mayor representación en el conjunto de prueba. Sin embargo, las clases Bajo y Alto mantienen valores altos de F1 (0.93 y 0.88), demostrando que:

- el balanceo aplicado (`class_weight='balanced'`),
- la normalización,
- y la selección de características

fueron adecuadas para asegurar un rendimiento equitativo entre clases con tamaños distintos. Además, la AUC-ROC de 0.99 confirma que el modelo no solo clasifica correctamente, sino que separa con claridad los distintos niveles de apropiación docente.

### 3.3 Análisis de Predictores Significativos

El análisis de importancia de características, utilizando Permutation Importance y la magnitud de los coeficientes de la MLR, permitió identificar cuáles variables aportaron más al proceso de predicción. Aunque se utilizó un total de 9 predictores, no todos contribuyeron con la misma fuerza. Los resultados muestran una jerarquía clara:

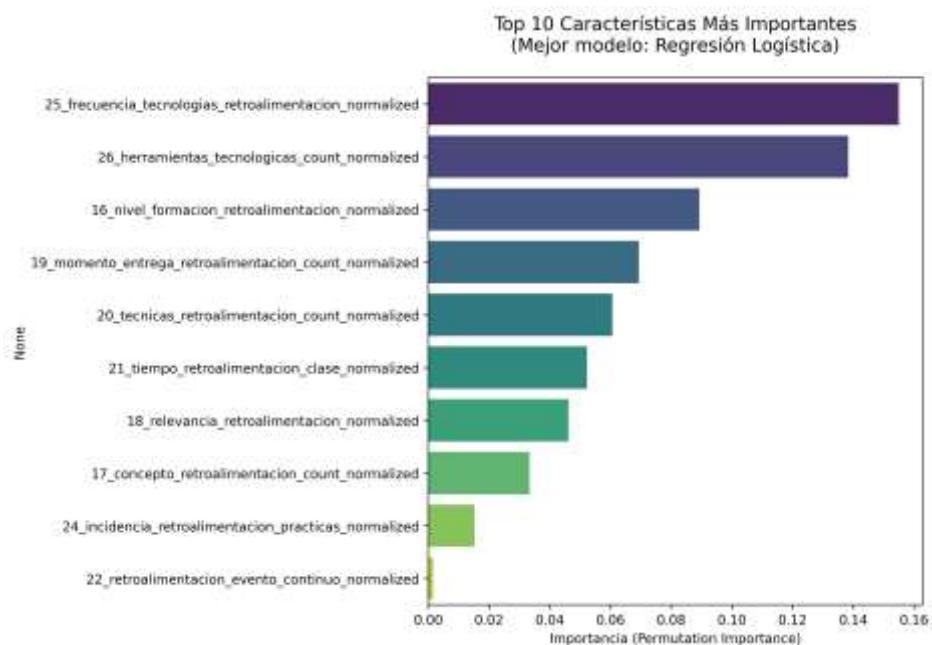
- **Frecuencia de Uso Tecnología y Herramientas Tecnológicas**

La frecuencia en el uso de tecnologías para entregar retroalimentación (plataformas, formularios, rúbricas digitales, asistentes de IA, etc.) fue el predictor más influyente y el uso de estas tecnologías en segundo lugar. Este resultado evidencia la creciente relación entre tecnologías educativas y prácticas de retroalimentación efectivas.

- **Nivel de Formación en Retroalimentación**

La formación formal o especializada (cursos, talleres, diplomados) aparece como el terce factor más determinante. Los docentes con mayor formación muestran patrones más consistentes con prácticas de retroalimentación alineadas al enfoque formativo.

Gráfico 1



### 3.4 Implicancias de los Hallazgos

Estos resultados permiten a las instituciones educativas orientar de manera eficiente sus esfuerzos de desarrollo profesional hacia los factores de mayor impacto:

- ampliar la oferta formativa especializada sobre retroalimentación;
- fomentar el uso estratégico de herramientas digitales como apoyo al proceso de retroalimentación.

En conjunto, estos predictores proporcionan un mapa claro para promover prácticas docentes que favorezcan niveles más altos de apropiación.

#### **4 Discusión**

Los hallazgos evidencian que es posible predecir de manera fiable el nivel de apropiación de la retroalimentación docente utilizando información autoinformada y técnicas de Machine Learning. El uso de la Regresión Logística Multinomial aportó interpretabilidad al modelo, permitiendo comprender el peso relativo de cada predictor y facilitando la generación de explicaciones comprensibles para instituciones educativas.

Sin embargo, el estudio presenta algunas limitaciones:

- a) Tamaño y diversidad muestral: Aunque la muestra es suficiente para entrenar el modelo, un mayor número de docentes de distintas regiones y tipos de establecimientos podría mejorar la generalización.
- b) Naturaleza autoinformada de los datos: Las respuestas podrían estar influidas por sesgos de deseabilidad social.
- c) Uso de un único tipo de modelo: Otros algoritmos, como Random Forest o Gradient Boosting, podrían explorarse para comparar desempeño y robustez.
- d) Predicción basada solo en cuestionarios: Incorporar datos observacionales, entrevistas o evaluaciones externas podrían incrementar la precisión del índice.

A futuro, se propone:

- Ampliar la muestra y actualizar el modelo con nuevas iteraciones.
- Integrar explicabilidad avanzada (SHAP, LIME) para análisis más profundo de los predictores.
- Explorar agentes más complejos que integren recomendaciones adaptativas basadas en el perfil del docente.
- Incorporar un módulo de seguimiento longitudinal para observar la progresión del nivel de apropiación.

#### **5 Conclusión**

La presente investigación demuestra que es posible diseñar, entrenar e implementar un agente inteligente capaz de predecir el nivel de apropiación de la retroalimentación docente mediante un modelo de Machine Learning integrado en una interfaz accesible. El índice construido permitió operacionalizar el constructo desde sus dimensiones teóricas, y el modelo predictivo mostró un desempeño adecuado y consistente.

LUMINA constituye una herramienta práctica, gratuita y escalable que puede ser adoptada por instituciones educativas para:

- 1) realizar diagnósticos masivos y rápidos;
- 2) focalizar los recursos de capacitación en los factores predictivos más fuertes (tecnología y formación); y
- 3) proveer a los docentes de un *feedback* inmediato y basado en sus propias percepciones y prácticas.

Los principales predictores identificados (formación y tecnología) sugieren que la apropiación efectiva de la retroalimentación no es solo una cuestión de técnica, sino fundamentalmente de preparación y acceso a uso de tecnologías.

En conjunto, los resultados validan el potencial del uso de inteligencia artificial en educación para apoyar procesos evaluativos y de mejora docente, abriendo oportunidades para futuras investigaciones que integren IA y desarrollo profesional en contextos escolares.

## Referencias

- Amaranti Pesce, M., & Contreras Pérez, G. (2025). Del feedback al feedforward: promoviendo la reflexión, participación y uso sostenible de la retroalimentación en futuros docentes en formación. *Páginas de Educación*, 18(1), e4454. <https://doi.org/10.22235/pe.v18i1.4454>.
- Baker, R. S., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In *Emerging Trends in the Social and Behavioral Sciences*. Wiley.
- Decreto 67. (2018). Reglamento de evaluación, calificación y promoción de alumnos. Ministerio de Educación, Chile.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>.
- Herrera Araya, D. (2023). Tesis por compendio: Retroalimentación en el desarrollo del pensamiento social en educación secundaria. Universidad Diego Portales / Universidad Alberto Hurtado.
- Hernandez Medina, P., Ramirez Torres, G., & Pinilla Rodriguez, D. (2024). Teaching credentials and academic performance: A panel data study in a Latin American School of Economics. *Revista Meta: Avaliação*, 16(51), e4259. <https://doi.org/10.22347/2175-2753v16i51.4259>
- Hwang, G. J., Xie, H., Wahab, M. N. A., & Bervell, B. (2020). A review of artificial intelligence in education (AIEd) trend: an evidence-based approach. *Interactive Learning Environments*, 1-17.
- Lira, A., Cortez, M., Sánchez, B. y Zett, I. (2021). Retroalimentación para el aprendizaje profundo. Nota Técnica. Líderes Educativos PUCV.
- Ríos-Muñoz, D., & Herrera-Araya, D. (2023). Retroalimentación docente y autoevaluación de los estudiantes en educación superior: innovación evaluativa en la formación inicial de profesores de educación primaria. *Revista Brasileira de Educação*, 28, e280123. <https://doi.org/10.1590/S1413-24782023280123>
- Silva de Carvalho Ramos, J., & Carvalho, C. F. (2021). Percepções de estudantes do ensino superior sobre o feedback docente e desempenho acadêmico. *Revista Brasileira de Educação*, 26, e260081. <https://doi.org/10.1590/S1413-24782021260081>.