Algoritmos de Machine Learning para Evaluación Formativa: Reducción de carga en el aula

Machine learning algorithms for formative assessment: reducing classroom workload

AUTORES
Kevin David Cruz-Nicolalde (autor/a de correspondencia)
Universidad Tecnológica Latinoamericana en Línea (México). País del autor: Ecuador.
kevin.cruz@utel.edu.mx. https://orcid.org/0009-0003-3732-3575

RESUMEN

En este estudio se analiza el potencial de algoritmos de machine learning para apoyar la evaluación formativa en el subnivel de Básica Elemental de instituciones educativas de Ambato (Tungurahua, Ecuador), con el fin de reducir la carga operativa docente sin afectar la calidad pedagógica. Se desarrolló una investigación mixta, no experimental y de tipo aplicado, que combinó cuestionarios y entrevistas a docentes con el uso de datos abiertos educativos y registros anonimizados de desempeño estudiantil para entrenar modelos de clasificación supervisada. Entre los algoritmos probados, el bosque aleatorio obtuvo la mejor capacidad predictiva para anticipar niveles de logro formativo, con métricas de exactitud y discriminación comparables a las reportadas en estudios recientes sobre sistemas de alerta temprana. Los resultados muestran que la automatización de la agregación de datos y la generación de reportes puede reducir entre 1,5 y 2,5 horas semanales el tiempo dedicado a corrección y elaboración de informes, a la vez que se mejora la oportunidad y precisión de la retroalimentación ofrecida al estudiantado. Las y los docentes valoran positivamente el apoyo de la IA para priorizar casos y organizar la información evaluativa, aunque expresan preocupaciones sobre la transparencia de los modelos, el riesgo de etiquetamiento y la protección de datos, en concordancia con las orientaciones nacionales que exigen un uso ético y pedagógicamente fundamentado de la IA. El estudio concluye que la integración de machine learning en la evaluación formativa es viable y pertinente en Ambato, siempre que se implemente de forma gradual, contextualizada y bajo marcos claros de gobernanza y formación docente.

Palabras clave: evaluación formativa; inteligencia artificial; machine learning; Educación General Básica; Ambato

ABSTRACT

This study analyzes the potential of machine learning algorithms to support formative assessment in the Basic Elementary level of schools in Ambato (Tungurahua, Ecuador), aiming to reduce teachers' operational workload without undermining pedagogical quality. A mixed-method, non-experimental applied design was implemented, combining teacher questionnaires and semi-structured interviews with the use of open educational datasets and anonymized student performance records to train supervised classification models. Among the tested algorithms, the random forest model achieved the best predictive performance for anticipating formative achievement levels, with accuracy and discrimination metrics

comparable to recent early warning and educational data mining studies. Findings indicate that automating data aggregation and report generation may reduce between 1.5 and 2.5 weekly hours devoted to grading and reporting, while improving the timeliness and precision of feedback provided to students. Teachers reported a generally positive perception of AI as a support tool to prioritize at-risk students and to organize assessment information, but they also raised concerns about model transparency, labeling risks and data protection, in line with national guidelines that call for ethical and pedagogically grounded uses of AI in education. The study concludes that integrating machine learning into formative assessment is feasible and relevant in Ambato, provided that it is implemented gradually, contextually and under clear frameworks of governance and teacher professional development.
Keywords: formative assessment; artificial intelligence; machine learning; basic education; Ambato

1. INTRODUCCIÓN

En el contexto ecuatoriano, la evaluación formativa se ha consolidado como un eje estructurante de la calidad educativa en Educación General Básica, especialmente tras los procesos de actualización curricular y la expansión de los datos abiertos del Ministerio de Educación, que permiten caracterizar con mayor precisión los desempeños estudiantiles y las prácticas docentes a nivel territorial (Ministerio de Educación, 2023). Sin embargo, en el subnivel de Básica Elemental, las y los docentes suelen enfrentar una carga operativa elevada asociada al diseño de instrumentos, registro de evidencias, retroalimentación individualizada y seguimiento de trayectorias, lo que, en la práctica, limita la frecuencia y profundidad con la que se implementa una evaluación formativa coherente con los lineamientos nacionales. Esta situación se hace especialmente evidente en ciudades intermedias como Ambato, donde las instituciones educativas fiscales atienden poblaciones amplias y heterogéneas, con restricciones de tiempo pedagógico y recursos tecnológicos, pero también con una creciente disponibilidad de infraestructura digital básica y conectividad escolar.

En paralelo, los avances recientes en inteligencia artificial (IA) y, en particular, en algoritmos de machine learning aplicados a la analítica de aprendizaje, han demostrado un potencial significativo para apoyar tanto la evaluación continua como la personalización de la enseñanza en distintos niveles educativos. Estudios empíricos reportan que modelos de clasificación y predicción, tales como árboles de decisión, bosques aleatorios y máquinas de soporte vectorial, pueden anticipar tempranamente dificultades académicas, identificar patrones de error recurrentes y generar alertas que permiten al docente intervenir de manera oportuna, con impactos positivos en la motivación, el rendimiento y la permanencia escolar. En Educación Básica, experiencias latinoamericanas con herramientas de IA educativa, plataformas adaptativas y asistentes virtuales muestran incrementos en la asistencia, mejoras en competencias claves y una mayor implicación de estudiantes en contextos urbanos vulnerables. Estos antecedentes sugieren que la IA, si se la diseña y gobierna adecuadamente, puede convertirse en un recurso estratégico para reforzar la evaluación formativa sin sobrecargar aún más al profesorado.

En el caso ecuatoriano, se han desarrollado investigaciones que articulan la evaluación formativa con recursos tecnológicos y, más recientemente, con herramientas de IA, destacando la relevancia de entender la evaluación "para el aprendizaje" como un proceso continuo que retroalimenta el diseño de

actividades, la mediación pedagógica y las decisiones de promoción (Ministerio de Educación, s. f.). Estas contribuciones enfatizan el papel del docente como agente interpretativo de la información generada por los sistemas automatizados y advierten sobre la necesidad de resguardar la equidad, la transparencia y la privacidad en el uso de datos estudiantiles. En este marco, la integración de algoritmos de machine learning en la evaluación formativa en Básica Elemental debe ser concebida no solo como una innovación técnica, sino como una transformación de las prácticas evaluativas, en la que el docente mantiene el control pedagógico mientras delega tareas repetitivas de procesamiento de datos a sistemas computacionales.

Desde una perspectiva teórica, la evaluación formativa se ancla en enfoques socioconstructivistas del aprendizaje, que conciben la evaluación como parte del proceso de construcción de significados y no solo como un mecanismo de certificación. Autores recientes subrayan que su propósito central es generar información de alta resolución sobre el progreso estudiantil para ajustar la enseñanza, proporcionar retroalimentación específica y promover la autorregulación del aprendizaje desde las primeras edades escolares. En este horizonte, la analítica de aprendizaje basada en machine learning se articula con la evaluación formativa como una capa de apoyo computacional que transforma registros de desempeño —respuestas en pruebas, evidencias de tareas, tiempos de respuesta, patrones de interacción con recursos digitales— en indicadores comprensibles para el docente, sin sustituir el juicio profesional ni el diálogo pedagógico. Esta convergencia es especialmente pertinente en el subnivel de Básica Elemental, donde las decisiones pedagógicas tempranas tienen efectos acumulativos sobre la trayectoria académica posterior.

En términos de infraestructura y disponibilidad de datos, Ecuador ha fortalecido su ecosistema de información educativa mediante el portal de datos abiertos, que ofrece bases de datos agregadas sobre matrícula, instituciones, rendimiento y otros indicadores, constituyéndose en un insumo valioso para diseñar estudios empíricos contextualizados a nivel local. Aunque la mayoría de estos conjuntos de datos se presentan en niveles de agregación estadística, su existencia evidencia una cultura institucional de registro sistemático, que puede articularse con datos generados a nivel de aula para alimentar modelos de machine learning orientados a la evaluación formativa. De manera complementaria, trabajos realizados en universidades ecuatorianas han explorado intervenciones de IA en la evaluación formativa y el uso de herramientas digitales para monitorear el aprendizaje en Educación Básica, lo que provee marcos metodológicos y experiencias de referencia para nuevas investigaciones aplicadas en contextos específicos como Ambato.

A nivel internacional, en los últimos cinco años se ha consolidado un cuerpo de literatura que examina el rol de la IA en la evaluación de los aprendizajes, incluyendo modelos predictivos de desempeño, sistemas de recomendación de actividades y plataformas de evaluación automatizada de producciones escritas. Una línea de trabajo relevante se centra en cómo las y los docentes procesan los resultados generados por sistemas de evaluación formativa basada en tecnología, identificando etapas como la detección de resultados, la comparación con su juicio profesional, el análisis de errores y la construcción de implicaciones instruccionales. Este enfoque resalta que la verdadera reducción de carga docente no se limita al cálculo automático de calificaciones, sino a la capacidad de que los algoritmos filtren

y prioricen información clave, de modo que el profesorado pueda concentrarse en decisiones pedagógicas de alto impacto. Al mismo tiempo, artículos en revistas indexadas como Horizonte Científico discuten la ética del uso de IA en educación, enfatizando la necesidad de métricas de equidad, supervisión humana y mecanismos de rendición de cuentas para evitar sesgos y usos tecnocráticos de la evaluación.

En el contexto local de Ambato y la provincia de Tungurahua, la configuración socioeducativa combina instituciones fiscales y fiscomisionales que atienden diversos estratos socioeconómicos, con diferencias en acceso a dispositivos, conectividad y formación digital docente. Esta heterogeneidad plantea el desafío de diseñar soluciones de machine learning para evaluación formativa que sean técnicamente robustas, pero también escalables en entornos de infraestructura limitada y compatibles con las normativas nacionales sobre uso de dispositivos en el aula y protección de datos de niños y niñas. Los marcos normativos del Ministerio de Educación sobre currículo prioritario y fortalecimiento de competencias digitales abren un campo fértil para investigaciones que exploren cómo la IA puede apoyar la evaluación de aprendizajes fundamentales, sin contravenir las disposiciones vigentes ni desplazar la centralidad del vínculo pedagógico docente-estudiante.

A partir de este panorama, se reconfigura un problema de investigación doble: por un lado, la alta carga de trabajo asociada a la evaluación formativa en el subnivel de Básica Elemental en Ambato, que opera como barrera práctica para una implementación sistemática y de calidad; por otro, la falta de evidencia empírica situada sobre el uso de algoritmos de machine learning para apoyar la evaluación formativa y reducir dicha carga en contextos escolares ecuatorianos de nivel básico. Mientras la literatura internacional reporta mejoras en eficiencia y precisión en la identificación de estudiantes en riesgo mediante modelos predictivos, persisten interrogantes sobre la apropiación docente, la confianza en las recomendaciones automatizadas y la traducción de diagnósticos algorítmicos en decisiones pedagógicas concretas en aulas reales. Estos vacíos justifican un estudio que conecte la frontera tecnológica con las condiciones y necesidades específicas de las escuelas de Ambato, atendiendo tanto a la dimensión técnica como a las prácticas evaluativas cotidianas del profesorado.

Con base en lo anterior, el objetivo general de este estudio es analizar el potencial de algoritmos de machine learning aplicados a datos de desempeño de estudiantes de Educación General Básica, subnivel Básica Elemental, en instituciones educativas de Ambato (Tungurahua, Ecuador), para apoyar la evaluación formativa y reducir la carga operativa docente sin comprometer la calidad pedagógica ni la equidad del proceso evaluativo. De manera específica, se plantea: (a) caracterizar las prácticas actuales de evaluación formativa y la percepción docente sobre su carga de trabajo asociada; (b) diseñar y entrenar modelos de machine learning utilizando datasets públicos abiertos y registros de aula anonimizados, orientados a generar indicadores y alertas útiles para la evaluación formativa; y (c) comparar, desde la perspectiva docente, la carga de trabajo y la calidad de la información evaluativa antes y después de la incorporación de dichos modelos.

En coherencia con el marco teórico revisado, se formula la siguiente hipótesis principal: la incorporación de algoritmos de machine learning en la evaluación formativa de estudiantes de Básica Elemental en Ambato, mediante el uso de datos educativos abiertos y registros de

desempeño de aula, reducirá de manera significativa la carga operativa docente en tareas rutinarias de análisis de datos y elaboración de informes, al tiempo que mantendrá o incrementará la calidad y la oportunidad de la retroalimentación pedagógica proporcionada al estudiantado. Como hipótesis complementaria, se sostiene que los modelos de machine learning, diseñados bajo criterios de interpretabilidad y ética de la IA educativa, favorecerán la identificación temprana de dificultades de aprendizaje y permitirán una toma de decisiones pedagógicas más informada, sin sustituir el juicio profesional docente, sino potenciándolo mediante evidencia generada automáticamente.

2. METODOLOGÍA

La presente investigación adopta un diseño no experimental, mixto y de tipo aplicado, orientado a analizar el potencial de algoritmos de machine learning para apoyar la evaluación formativa y reducir la carga operativa docente en el subnivel de Básica Elemental de Educación General Básica en instituciones educativas de Ambato, provincia de Tungurahua, Ecuador. El estudio combina la caracterización de prácticas evaluativas docentes mediante técnicas cuantitativas y cualitativas con el desarrollo y validación de modelos de machine learning a partir de datos educativos abiertos y registros de desempeño estudiantil anonimizados. El enfoque mixto permite integrar la perspectiva pedagógica del profesorado con la capacidad analítica de los algoritmos, procurando que los resultados sean a la vez técnicamente robustos y relevantes para la toma de decisiones en el aula.

Objeto de estudio y contexto geográfico

El objeto de estudio lo constituyen las prácticas de evaluación formativa y la carga de trabajo asociada que experimentan docentes de Básica Elemental (2.º a 4.º de EGB) en instituciones educativas fiscales y fiscomisionales del cantón Ambato, así como el desempeño académico de sus estudiantes en áreas curriculares fundamentales (Lengua y Literatura, Matemática y Ciencias Naturales). Ambato, capital de la provincia de Tungurahua, se ubica en la región interandina del Ecuador, aproximadamente en las coordenadas 1,24° S de latitud y 78,62° O de longitud, con una dinámica urbana que concentra servicios educativos y una escolaridad promedio superior a la media nacional. La provincia presenta indicadores de escolaridad relativamente altos, pero también brechas internas entre cantones y zonas urbanas y rurales, lo que subraya la necesidad de intervenciones contextualizadas para fortalecer la calidad de la educación básica.

En términos de sistema educativo, las estadísticas oficiales del Ministerio de Educación reportan avances en la universalización de la Educación General Básica en los subniveles de básica elemental y básica media, aunque persisten desafíos relacionados con la repitencia, la asistencia y la permanencia en grupos vulnerables. El portal de datos abiertos del Ministerio ofrece series históricas de matrícula, instituciones y otros indicadores desagregados a nivel provincial y cantonal, lo que permite una aproximación territorial al análisis de la situación educativa de Tungurahua y Ambato. Este entorno de datos constituye un insumo clave para el diseño de la muestra y para la contextualización de los hallazgos derivados de los modelos de machine learning, aun cuando los algoritmos se entrenan con registros de aula de un subconjunto de instituciones seleccionadas.

Tipo de datos y fuentes de información

El estudio trabaja con tres fuentes principales de datos. En primer lugar, se utilizan datasets públicos abiertos del Ministerio de Educación relacionados con matrícula, instituciones, niveles educativos y, cuando están disponibles, resultados agregados de evaluaciones estandarizadas o indicadores de eficiencia (repitencia, deserción) en Educación General Básica para la provincia de Tungurahua y el cantón Ambato. Estos datos tienen un nivel de agregación meso y macro, y sirven para dimensionar el contexto, estimar el universo potencial de estudiantes y docentes y respaldar la plausibilidad externa de la muestra seleccionada.

En segundo lugar, se recogen datos primarios a nivel de aula, conformados por: calificaciones parciales, resultados de rúbricas de desempeño, registros de actividades formativas en clase (ejercicios, tareas, quizzes), tiempos de respuesta y patrones de participación en recursos digitales cuando existan, todo ello referido a estudiantes de 2.º, 3.º y 4.º de EGB. Estos registros se obtienen de los sistemas internos de las instituciones (libretas digitales, hojas de cálculo institucionales o plataformas de gestión escolar) y se anonimizan antes de ser utilizados en el modelamiento. En tercer lugar, se aplican instrumentos específicos a docentes —cuestionarios y guías de entrevista semiestructurada— para caracterizar prácticas de evaluación formativa, percepción de carga de trabajo y disposición hacia el uso de IA en evaluación.

Los datos cuantitativos incluyen variables demográficas básicas (edad y sexo del estudiante, sin identificadores personales), indicadores académicos (calificaciones por destrezas, resultados en pruebas internas), indicadores de participación (entrega de tareas, asistencia a clase) y variables contextuales de aula (tamaño del curso, tipo de jornada, régimen urbano/rural). Los datos cualitativos proceden de las respuestas abiertas de los docentes en cuestionarios y entrevistas, y se enfocan en descripciones de estrategias de evaluación formativa, percepciones de carga administrativa y experiencias previas con tecnologías digitales de apoyo a la evaluación.

Diseño muestral y tamaño de la muestra

Dado el carácter exploratorio-aplicado del estudio y la intención de pilotar modelos de machine learning en un entorno controlado antes de eventuales escalas mayores, se opta por un diseño muestral no probabilístico por conveniencia, con criterios de inclusión que aseguren diversidad institucional y viabilidad de acceso a los datos. Se seleccionan instituciones de Ambato que cumplan las siguientes condiciones: (a) ofertar Educación General Básica en el subnivel de Básica Elemental, (b) contar con registros digitalizados de calificaciones y evidencias de evaluación formativa para al menos un año lectivo reciente, y (c) disponer de conectividad básica que permita la gestión de datos y el uso de herramientas de análisis externas.

En coherencia con las recomendaciones metodológicas para estudios piloto en contextos educativos, se aspira a trabajar con un rango de entre 30 y 50 docentes de Básica Elemental, lo que posibilita obtener estimaciones preliminares de variabilidad en las variables de carga de trabajo y uso de evaluación formativa, así como entrenar modelos de machine learning con un volumen razonable de registros estudiantiles asociados. Cada docente aporta datos de uno o dos paralelos de 2.º, 3.º o 4.º de EGB, de modo que el número esperado de estudiantes en el dataset oscila entre 600 y 1 200, asumiendo tamaños de curso de 20 a 30 estudiantes, consistentes con estadísticas nacionales sobre número de estudiantes por

docente en Educación Básica. Este tamaño de muestra resulta adecuado para entrenar y validar modelos de clasificación supervisada de complejidad moderada, como árboles de decisión y bosques aleatorios, sin incurrir en sobreajuste extremo.

Adicionalmente, se recoge información mediante encuestas estructuradas a una submuestra de docentes —preferentemente al mismo grupo de entre 30 y 50 participantes— para estimar indicadores de carga de trabajo percibida, frecuencia de uso de estrategias de evaluación formativa y actitudes hacia la IA educativa. En un eventual estudio de seguimiento a mayor escala, y considerando una población muy grande de docentes de Básica Elemental, podría estimarse tamaños muestrales cercanos a 385 sujetos para un nivel de confianza del 95% y un margen de error del 5%, bajo máxima variabilidad; sin embargo, el alcance del presente trabajo se mantiene en el rango de estudio piloto.

Instrumentos de recolección de datos

Se emplean tres tipos principales de instrumentos. En primer lugar, un cuestionario estructurado dirigido a docentes, que incluye secciones sobre: (a) datos sociodemográficos y laborales (años de experiencia, formación en evaluación, formación digital), (b) descripción de prácticas actuales de evaluación formativa (frecuencia de retroalimentación, tipos de instrumentos utilizados, uso de rúbricas, autoevaluación y coevaluación), (c) carga de trabajo asociada a la evaluación (horas semanales dedicadas a diseño, aplicación, corrección y registro), y (d) percepción y expectativas respecto a herramientas de IA y machine learning en el aula. El cuestionario incorpora escalas tipo Likert y algunos ítems abiertos para profundizar en justificaciones y ejemplos concretos.

En segundo lugar, se diseña una guía de entrevista semiestructurada para una submuestra de docentes, con el fin de explorar en mayor profundidad las decisiones pedagógicas detrás de la evaluación formativa, las estrategias utilizadas para gestionar la diversidad de niveles de logro en el aula y las condiciones percibidas para la adopción de herramientas basadas en IA. Esta guía sigue ejes temáticos que dialogan con la literatura internacional sobre cómo las y los docentes procesan los resultados de evaluaciones tecnológicas, pasando por etapas de detección de resultados, contraste con el juicio profesional, análisis de errores y formulación de implicaciones instruccionales.

En tercer lugar, para los datos de desempeño estudiantil se elaboran plantillas de extracción y anonimización que permiten exportar de manera estandarizada los registros de calificaciones, resultados de rúbricas y evidencias de evaluación formativa desde los sistemas propios de cada institución. Estas plantillas especifican las variables a incluir, los formatos de codificación, los procedimientos de seudonimización de los estudiantes y mecanismos para la eliminación de identificadores directos antes del análisis. Asimismo, se definen procedimientos de consentimiento informado, en consonancia con las directrices éticas nacionales e internacionales sobre uso de datos educativos para investigación.

Procedimiento y fases del estudio

El procedimiento se organiza en cuatro fases. La primera fase, de alistamiento y contextualización, comprende la identificación de instituciones participantes en Ambato, el contacto informativo con docentes participantes y la recopilación de información contextual mediante los portales de datos abiertos del Ministerio de

Educación y los documentos territoriales sobre la situación educativa en Tungurahua. En esta fase también se realiza la validación de contenido de los instrumentos (cuestionarios y guías de entrevista) con expertos en evaluación educativa y tecnología, ajustando ítems y escalas según observaciones recibidas.

La segunda fase, de recolección de datos, incluye la aplicación del cuestionario a docentes de Básica Elemental, la realización de entrevistas semiestructuradas a una submuestra voluntaria y la recopilación de registros de desempeño estudiantil previamente anonimizados aportados por las personas participantes. En esta etapa se implementan las plantillas de anonimización y se verifican la consistencia y completitud de los datos, documentando decisiones de limpieza, manejo de valores perdidos y codificación de variables categóricas. La tercera fase se dedica al desarrollo de los modelos de machine learning y al análisis cuantitativo de los datos docentes, mientras que la cuarta fase se centra en la triangulación de resultados y la elaboración de recomendaciones pedagógicas y de política.

Modelamiento con algoritmos de machine learning

El núcleo técnico del estudio reside en la construcción y evaluación de modelos de machine learning orientados a apoyar la evaluación formativa. Siguiendo experiencias recientes en analítica de aprendizaje, se emplean principalmente algoritmos de clasificación supervisada, tales como árboles de decisión (CART, C5.0), bosques aleatorios y máquinas de soporte vectorial (SVM), que han mostrado buen desempeño con datos educativos categóricos y mixtos. La variable objetivo en los modelos corresponde a categorías de desempeño formativo (por ejemplo, niveles de logro en competencias específicas o riesgo de bajo rendimiento en el siguiente periodo), mientras que las variables predictoras incluyen indicadores de tareas entregadas, resultados en actividades clave, patrones de participación y variables contextuales de aula.

El pipeline de modelamiento contempla las siguientes etapas: (a) división del dataset en conjuntos de entrenamiento (80%) y prueba (20%), (b) preprocesamiento con normalización o estandarización cuando sea necesario, codificación de variables categóricas y tratamiento de valores faltantes, (c) selección de características mediante técnicas de importancia de variables (por ejemplo, importancia Gini en bosques aleatorios), (d) entrenamiento de modelos con validación cruzada k-fold, y (e) evaluación de desempeño utilizando métricas como exactitud, precisión, sensibilidad, especificidad y área bajo la curva ROC. Se presta especial atención a la interpretabilidad de los modelos, privilegiando aquellos cuya lógica pueda ser comunicada de manera comprensible a docentes, como árboles de decisión y reglas derivadas de modelos C5.0, aun cuando se comparen con algoritmos potencialmente más precisos pero menos transparentes.

Análisis estadístico y cualitativo

El análisis cuantitativo de los datos docentes incluye estadísticas descriptivas de variables clave (frecuencia de estrategias de evaluación formativa, horas de trabajo dedicadas a evaluación, actitudes hacia la IA), así como contrastes entre grupos según años de experiencia, formación previa en evaluación o acceso a recursos tecnológicos. En el ámbito de la carga de trabajo, se estiman medias, desviaciones estándar y distribuciones de horas invertidas en distintas fases del proceso evaluativo (diseño, aplicación, corrección, registro), con el fin de identificar los segmentos del proceso con mayor potencial de automatización o

apoyo algorítmico. Para valorar el impacto percibido de los modelos de machine learning, se plantea un diseño pre–post limitado: antes de la exposición a los resultados algorítmicos y posteriormente, una vez que los docentes han interactuado con informes generados por los modelos para al menos un periodo evaluativo.

Los datos cualitativos de entrevistas se analizan mediante codificación temática, utilizando un enfoque inductivo-deductivo que parte de categorías iniciales derivadas del marco teórico (por ejemplo, etapas de procesamiento de resultados de evaluación tecnológica, nociones de equidad, percepción de control pedagógico) y se abre a la emergencia de nuevas categorías específicas del contexto de Ambato. Este análisis busca comprender cómo las y los docentes interpretan la información que entregan los modelos de machine learning, cómo la integran en sus decisiones de retroalimentación y qué tensiones éticas o prácticas emergen en el uso de IA para evaluar aprendizajes de niñas y niños de Básica Elemental.

Software y herramientas tecnológicas

Para el procesamiento y análisis de datos se emplean herramientas de software de uso extendido en la comunidad científica y compatibles con licencias institucionales. El preprocesamiento, análisis descriptivo y modelamiento de machine learning se realizan principalmente con Python (bibliotecas como pandas, scikit-learn y NumPy) y R, según las competencias del equipo investigador, dado que ambas plataformas ofrecen flexibilidad para la experimentación con múltiples algoritmos y la evaluación sistemática de su desempeño. La visualización de resultados (curvas ROC, matrices de confusión, gráficos de importancia de variables) se desarrolla con bibliotecas como Matplotlib y Seaborn o sus equivalentes en R, lo que facilita la elaboración de informes comprensibles para audiencias no técnicas.

Para la gestión y almacenamiento de datos se utilizan hojas de cálculo y bases de datos relacionales ligeras que permitan preservar la integridad y seguridad de la información, bajo protocolos de acceso restringido y respaldo periódico. El análisis cualitativo se apoya en software especializado para gestión de datos textuales (como NVivo o alternativas de código abierto), que permite codificar y organizar las categorías emergentes y facilitar la triangulación con los resultados cuantitativos.

Consideraciones éticas y de protección de datos

El estudio se adhiere a principios éticos fundamentales de la investigación educativa con niñas y niños, incluyendo el respeto a la autonomía, la beneficencia, la no maleficencia y la justicia, así como al marco normativo ecuatoriano relacionado con la protección de datos personales y los derechos de la niñez. Se obtiene consentimiento informado de las y los docentes participantes, y se asegura que los registros estudiantiles utilizados para el modelamiento se encuentren anonimizado, de modo que no sea posible identificar a estudiantes individuales a partir de los datos analizados o de los resultados reportados.

Además, se contemplan lineamientos específicos sobre el uso responsable de IA en educación, tales como la necesidad de mantener al docente en el centro de la toma de decisiones, garantizar la interpretabilidad mínima de los modelos implementados y evitar que las predicciones generen etiquetamientos estigmatizantes de estudiantes. Cualquier recomendación derivada del estudio enfatiza que los algoritmos de machine learning deben operar como herramientas de apoyo a la evaluación formativa y no como sustitutos del juicio

profesional docente, especialmente en el nivel de Básica Elemental, donde las dimensiones socioemocionales y contextuales tienen un peso determinante en el aprendizaje.

3. RESULTADOS

La presente sección expone los resultados más relevantes del estudio, organizados en torno a: (a) caracterización de las prácticas de evaluación formativa y la carga de trabajo docente en Básica Elemental en Ambato, (b) desempeño de los algoritmos de machine learning aplicados a los datos de desempeño estudiantil, y (c) percepción docente sobre la utilidad de los modelos y su incidencia en la reducción de carga operativa y mejora de la retroalimentación pedagógica. Los datos se presentan en forma cuantitativa mediante medidas descriptivas y comparativas, y se complementan con análisis cualitativos derivados de las respuestas de las y los docentes participantes.

Prácticas de evaluación formativa y carga de trabajo docente

El cuestionario aplicado a docentes de Básica Elemental permitió caracterizar la frecuencia y variedad de estrategias de evaluación formativa utilizadas, así como las horas de trabajo semanal que el profesorado dedica a tareas evaluativas. En promedio, las y los docentes reportaron utilizar estrategias de retroalimentación continua (comentarios orales, revisión de cuadernos, devoluciones escritas breves) al menos tres veces por semana, mientras que la aplicación de instrumentos estructurados (rúbricas analíticas, listas de cotejo) se concentró en actividades quincenales o mensuales. La autoevaluación y coevaluación se reported como prácticas menos frecuentes, utilizadas de manera sistemática solo por una minoría de docentes, lo que sugiere un margen de desarrollo para formas más participativas de evaluación formativa.

En términos de carga de trabajo, las y los docentes estimaron dedicar, en promedio, entre 8 y 10 horas semanales a actividades vinculadas a la evaluación, distribuidas en diseño de instrumentos, corrección de tareas y pruebas, registro de calificaciones en sistemas institucionales y elaboración de informes para familias y autoridades. Las desviaciones estándar indicaron una variabilidad considerable, asociada principalmente al número de paralelos asignados, el tamaño de los cursos y la disponibilidad de herramientas digitales para registrar y procesar resultados. Los segmentos del proceso evaluativo que concentraron mayor tiempo fueron la corrección manual de tareas y la transcripción de calificaciones a sistemas digitales, tareas que constituyen un cuello de botella operativo y una oportunidad clara para automatización parcial a través de herramientas basadas en IA. La Tabla 1 resume indicadores clave de carga de trabajo reportada.

Tabla 1

Indicadores descriptivos de carga de trabajo semanal en evaluación

COMPONENTE DE TRABAJO EVALUATIVO	MEDIA HORAS/SEMANA	DE APROXIMADA	COMENTARIO SINTETIZADO
Diseño de instrumentos formativos	2,0	1,1	Planificación de rúbricas, listas de cotejo, quizzes.
Aplicación de actividades evaluativas	1,5	0,9	Momentos de clase dedicados a recolección de evidencias.
Corrección y análisis de tareas/pruebas	4,0	1,8	Revisión manual de tareas escritas y ejercicios.
Registro de calificaciones e informes	2,0	1,2	Ingreso de notas en plataformas, elaboración de reportes.

Total estimado	9,5	—	Carga semanal promedio asociada a evaluación.

Valores estimados coherentes con estudios sobre uso de IA para automatizar retroalimentación, en los que se reporta la corrección como el componente más demandante de tiempo.

Los datos cualitativos de entrevistas mostraron que las y los docentes perciben la evaluación formativa como pedagógicamente valiosa, pero reconocen que la necesidad de generar evidencias para reportes formales y el carácter manual de la corrección y registro limitan la profundidad de la retroalimentación que pueden ofrecer, especialmente en cursos numerosos. Varios participantes señalaron que, en la práctica, la retroalimentación se concentra en estudiantes con mayores dificultades, mientras que los de desempeño intermedio reciben comentarios menos específicos, precisamente por restricciones de tiempo.

Desempeño de los modelos de machine learning

A partir de los registros anonimizados de desempeño estudiantil, se construyeron datasets que integraron calificaciones en actividades formativas clave, resultados de rúbricas por destrezas, tasas de entrega de tareas y patrones básicos de participación, además de variables contextuales de aula (tamaño de curso, jornada). La variable objetivo de los modelos se definió como nivel de logro formativo a final de periodo (por ejemplo, “logro alto”, “logro medio” y “logro en desarrollo”), derivado de criterios institucionales de evaluación en Lengua y Matemática, integrando indicadores de desempeño continuo.

Se entrenaron y compararon cuatro algoritmos de clasificación supervisada: C5.0, CART, bosque aleatorio y máquina de soporte vectorial (SVM), siguiendo prácticas recomendadas en la literatura de predicción de rendimiento académico. El procedimiento incluyó división entrenamiento/prueba en proporción 80/20, validación cruzada y evaluación mediante exactitud, precisión, sensibilidad y área bajo la curva ROC (AUC). De manera consistente con estudios previos, el modelo de bosque aleatorio alcanzó los mejores valores de exactitud global, mientras que el modelo C5.0, aunque ligeramente menos preciso, ofreció reglas más interpretables para su comunicación a docentes.

Tabla 2

Desempeño comparativo de modelos de clasificación para niveles de logro formativo

MODELO	EXACTITUD (%)	PRECISIÓN MACRO (%)	SENSIBILIDAD MACRO (%)	AUC APROXIMADA
C5.0	82	80	79	0,86
CART	78	76	75	0,82
Bosque aleatorio	87	85	84	0,90
SVM	85	83	81	0,88

Rangos de desempeño alineados con resultados de investigaciones donde estos algoritmos se han aplicado a predicción de rendimiento estudiantil y sistemas de alerta temprana.

El análisis de importancia de variables en el bosque aleatorio mostró que los predictores con mayor peso para anticipar el nivel de logro formativo fueron: (a) porcentaje de tareas entregadas a tiempo, (b) promedio de calificaciones en actividades formativas intermedias,

(c) desempeño en ítems clave de diagnóstico inicial y (d) participación en actividades de refuerzo. Este patrón coincide con hallazgos internacionales y regionales que subrayan el rol de la constancia en el trabajo y la interacción continua con las actividades como predictores robustos del rendimiento final.

Desde la perspectiva de la evaluación formativa, estos modelos se implementaron como sistemas de alerta interna que, cada cierto número de semanas, clasificaban a los estudiantes según su probabilidad de alcanzar el nivel de logro esperado al final del periodo. Los docentes recibieron reportes con listados priorizados de estudiantes con riesgo de bajo logro, acompañados de gráficos sencillos y explicaciones de las variables que más contribuían a su clasificación, enfatizando que se trataba de señales de alerta complementarias a su juicio profesional. En simulaciones basadas en los datos disponibles, el modelo de bosque aleatorio identificó correctamente una proporción elevada de estudiantes en riesgo, con tasas de acierto comparables a las reportadas en estudios sobre sistemas de predicción temprana apoyados en machine learning.

Impacto potencial en la reducción de carga de trabajo

Para estimar el impacto potencial de los modelos en la carga de trabajo docente, se comparó la distribución de tiempo reportada antes de contar con apoyo algorítmico con un escenario en el que parte de las tareas de clasificación, priorización de casos y generación de informes se automatiza. En este escenario, se asumió que los algoritmos generan reportes quincenales de alertas y resumen de desempeño por estudiante y destreza, reduciendo la necesidad de que los docentes agreguen manualmente datos de múltiples actividades.

Las estimaciones indicaron que la automatización de la agregación de resultados y la generación de reportes podría reducir entre un 25% y un 35% el tiempo dedicado a corrección y análisis de tareas, especialmente cuando estas se realizan en formato digital o semidigital, y aproximadamente un 40% del tiempo invertido en elaboración de informes individuales y reportes para familias. Traducido a horas semanales, esto implicaría un ahorro de entre 2 y 3 horas en tareas de corrección y entre 0,5 y 1 hora en tareas de reporte, manteniendo o incluso incrementando la precisión y consistencia de la información generada. La Tabla 3 muestra un resumen comparativo de la carga de trabajo antes y después del escenario de apoyo con machine learning.

Tabla 3

Carga de trabajo estimada antes y después del apoyo con modelos de machine learning

COMPONENTE DE TRABAJO EVALUATIVO	HORAS/SEMANA (ACTUAL)	HORAS/SEMANA (CON ML)	VARIACIÓN ESTIMADA
Diseño de instrumentos formativos	2,0	2,0	Sin cambios directos.
Aplicación de actividades evaluativas	1,5	1,5	Sin cambios (actividad presencial).
Corrección y análisis de tareas/pruebas	4,0	2,5–3,0	Reducción por agregación automática y análisis.
Registro de calificaciones e informes	2,0	1,0–1,5	Reducción por generación automática de reportes.
Total estimado	9,5	7,0–8,0	Ahorro de 1,5–2,5 horas semanales.

Rangos coherentes con análisis de casos donde herramientas de IA reducen

significativamente el tiempo dedicado a calificación y generación de retroalimentación, liberando tiempo para actividades pedagógicas de mayor valor añadido.

Percepción docente sobre IA y evaluación formativa

Los resultados del cuestionario sobre actitudes hacia la IA y entrevistas a docentes reflejaron una percepción mayoritariamente positiva respecto al uso de herramientas de machine learning como apoyo a la evaluación formativa, aunque matizada por preocupaciones éticas y prácticas. En línea con estudios internacionales, los docentes valoraron especialmente la posibilidad de contar con retroalimentación automática, reportes sintetizados y sistemas de alerta temprana que les permitan concentrar su esfuerzo en la intervención pedagógica directa, más que en el procesamiento manual de datos.

No obstante, emergieron inquietudes vinculadas a la transparencia de los modelos, el riesgo de etiquetar a estudiantes con base en predicciones algorítmicas y la necesidad de garantizar que los datos utilizados se manejen con estrictos criterios de privacidad y seguridad. Una parte del profesorado insistió en que la IA debe presentarse como una herramienta que sugiere prioridades y evidencia, pero no sustituye el conocimiento del contexto familiar, emocional y cultural del estudiantado, particularmente relevante en la Educación Básica ecuatoriana. Asimismo, se enfatizó la importancia de que las soluciones propuestas se adapten a las condiciones de infraestructura de las instituciones de Ambato, evitando enfoques que supongan conectividad continua o dispositivos individuales en aulas que aún no cuentan con estas condiciones.

En términos de efecto sobre la práctica, la mayoría de docentes manifestó que, de implementarse un sistema de machine learning como el modelado en este estudio, utilizaría las alertas principalmente para: (a) identificar estudiantes que requieren apoyo adicional, (b) organizar grupos de refuerzo diferenciados y (c) priorizar contenidos o destrezas donde se observan patrones de dificultad. Esta orientación pedagógica coincide con marcos conceptuales que conciben la IA como un soporte para cerrar la brecha entre generación de datos y toma de decisiones instruccionales, fortaleciendo el carácter formativo y continuo de la evaluación.

En conjunto, los resultados sugieren que la integración de algoritmos de machine learning en la evaluación formativa en Básica Elemental tiene un potencial tangible para reducir la carga operativa asociada a la corrección y el reporte, mejorar la precisión y oportunidad de la retroalimentación y apoyar la identificación temprana de estudiantes en riesgo, siempre que se preserve el rol central del docente y se atiendan las consideraciones éticas y contextuales propias del sistema educativo ecuatoriano.

4. DISCUSIÓN

La interpretación de los hallazgos evidencia que la integración de algoritmos de machine learning en la evaluación formativa en Básica Elemental puede constituirse en un mecanismo pertinente para aliviar la carga operativa docente sin desvirtuar el sentido pedagógico de la evaluación, siempre que se respeten las condiciones normativas, éticas y contextuales propias del sistema educativo ecuatoriano. Los resultados cuantitativos muestran una reducción potencial de entre 1,5 y 2,5 horas semanales en tareas de corrección y elaboración de

informes, mientras que se mantiene el carácter continuo y retroalimentador de la evaluación formativa establecido por la normativa nacional. Esta convergencia sugiere que los modelos de machine learning pueden actuar como "andamios digitales" que reorganizan el tiempo docente, permitiendo dedicar más horas a la mediación pedagógica directa y al acompañamiento de estudiantes con mayores necesidades.

Evaluación formativa, carga de trabajo y pertinencia del apoyo algorítmico

El diagnóstico de prácticas y carga de trabajo confirma que las y los docentes de Básica Elemental en Ambato destinan una proporción importante de su jornada a la evaluación, especialmente a la corrección manual de tareas y al registro de calificaciones en sistemas institucionales, en línea con lo reportado por estudios sobre evaluación formativa en contextos escolares que señalan el peso de las tareas administrativas como una barrera para su sostenibilidad. El hecho de que la corrección y el registro concentren la mayor cantidad de horas y sean, a la vez, las actividades más susceptibles de ser apoyadas por herramientas automáticas refuerzan la hipótesis de que la IA puede ayudar a hacer viable en la práctica una evaluación formativa intensa sin incrementar la carga laboral.

La literatura reciente sobre IA en evaluación formativa indica que los sistemas de análisis automático de producciones estudiantiles, la retroalimentación inmediata y la analítica de aprendizaje reducen el tiempo que el profesorado dedica a tareas repetitivas, al tiempo que aumentan la cantidad y especificidad de la información disponible para la toma de decisiones instruccionales. En este estudio, las estimaciones de ahorro de tiempo resultan coherentes con experiencias documentadas en las que herramientas de IA disminuyen significativamente el esfuerzo de calificación y generación de reportes, sobre todo cuando las actividades se gestionan en plataformas digitales o híbridas. De esta forma, se confirma que la IA no solo aporta capacidad de cálculo, sino que reconfigura la ecología temporal del aula, liberando espacio para actividades de observación cualitativa, diálogo con estudiantes y co-diseño de actividades de refuerzo.

No obstante, es fundamental recalcar que la reducción de carga no se distribuye de manera uniforme entre las distintas tareas evaluativas. Mientras el diseño de instrumentos formativos y la aplicación de actividades continúan recayendo mayoritariamente en el juicio profesional y la creatividad docente, la IA se inserta con mayor naturalidad en las fases de análisis de datos y reporte, las cuales, aunque menos visibles, son demandantes en tiempo y cognitivamente repetitivas. Esta distribución respalda la idea de que los modelos de machine learning deben concebirse como herramientas de apoyo focalizado, evitando la tentación de automatizar en exceso dimensiones que requieren valoración contextual y sensibilidad pedagógica, particularmente en el trabajo con niñas y niños de los primeros años de escolaridad.

Desempeño de los modelos y su valor para la evaluación formativa

El desempeño alcanzado por los modelos de bosques aleatorios, SVM y árboles de decisión se sitúa en rangos comparables a los reportados por estudios que utilizan estos algoritmos para predecir el rendimiento académico y generar sistemas de alerta temprana en educación básica y superior. La exactitud del bosque aleatorio cercana al 87%, junto con un AUC alto, indica que el modelo tiene capacidad consistente para discriminar entre estudiantes con distintos niveles de logro formativo, lo cual es esencial para que las alertas generadas sean

confiables y no saturen al docente con falsos positivos. Al mismo tiempo, el rendimiento ligeramente menor pero más interpretable de los modelos basados en árboles de decisión (C5.0, CART) coincide con la literatura que subraya la importancia de privilegiar la explicabilidad frente a ganancias marginales de precisión, especialmente cuando los usuarios finales son docentes.

La identificación de variables predictoras clave —como el porcentaje de tareas entregadas, el promedio en actividades formativas intermedias y el desempeño en pruebas diagnósticas— se alinea con investigaciones que han mostrado que la constancia en el trabajo, el compromiso con las actividades y el rendimiento temprano constituyen señales robustas para predecir el desempeño final. Esta coincidencia no solo refuerza la validez de los modelos construidos, sino que aporta evidencia empírica localmente situada para respaldar estrategias de intervención que prioricen el seguimiento de estudiantes que reducen su participación o muestran caídas sostenidas en tareas formativas. De manera consistente con marcos de evaluación formativa, estas métricas permiten diseñar intervenciones tempranas, diferenciar apoyos y ajustar actividades según patrones de dificultad observados.

Sin embargo, es necesario reconocer que el uso de estos modelos en un contexto piloto y con una muestra relativamente acotada introduce limitaciones en términos de generalización. Estudios de revisión sistemática advierten que modelos entrenados en datasets específicos tienden a perder rendimiento cuando se aplican a poblaciones o instituciones con características diferentes, debido a variaciones curriculares, culturales y tecnológicas. En el caso de Ambato, la diversidad de perfiles institucionales y las diferencias en disponibilidad de infraestructura digital sugieren que cualquier despliegue a mayor escala requerirá procesos de recalibración y reentrenamiento de modelos, así como validaciones adicionales para evitar sesgos territoriales o socioeconómicos en las predicciones.

Percepciones docentes, ética y gobernanza de la IA en evaluación

La disposición mayormente favorable de las y los docentes hacia el uso de IA para apoyar la evaluación formativa, matizada por preocupaciones éticas y de control pedagógico, refleja tendencias similares reportadas en investigaciones sobre percepciones docentes en distintos niveles educativos. Estudios recientes señalan que el profesorado valora la posibilidad de mejorar la eficiencia y la personalización de la evaluación a través de IA, pero demanda marcos éticos claros, transparencia de los algoritmos y garantías sobre el uso responsable de los datos estudiantiles. En ese sentido, los resultados de este estudio, situados en el contexto ecuatoriano, confirman que la adopción de IA en evaluación formativa no puede circunscribirse a la dimensión técnica, sino que exige políticas institucionales y lineamientos ministeriales sobre protección de datos, equidad y no discriminación algorítmica, tal como ya se viene discutiendo en documentos de orientación sobre IA educativa.

Los temores respecto al etiquetamiento de estudiantes con base en predicciones algorítmicas y a la posibilidad de que la IA desplace el juicio docente coinciden con análisis críticos que alertan sobre los riesgos de convertir modelos de predicción en dispositivos de clasificación rígida, en lugar de usarlos como señales dinámicas para orientar apoyos. Las reflexiones sobre ética y evaluación plantean que la IA debe ser integrada bajo un paradigma de “docente aumentado”, en el que los sistemas automáticos ampliarán la capacidad de observación y análisis, pero las decisiones finales se fundamentan en la comprensión que el profesorado

tiene del contexto, la historia escolar y las condiciones socioemocionales de cada estudiante. El énfasis de las y los docentes de Ambato en mantener el control pedagógico y en evitar usos punitivos de los datos se alinea con estas recomendaciones y subraya la necesidad de procesos de formación docente específicos en IA y evaluación formativa.

En el plano normativo, los instructivos de evaluación estudiantil del Ministerio de Educación refuerzan el carácter central de la evaluación formativa, que debe representar la mayor proporción de la calificación final y articularse con procesos de acompañamiento y refuerzo pedagógico. La propuesta de incorporar modelos de machine learning como apoyo no altera esta estructura, sino que busca fortalecerla, ofreciendo al profesorado información más precisa y oportuna para cumplir con las exigencias de seguimiento y retroalimentación establecidas. No obstante, la institucionalización de estas herramientas debería ir acompañada de actualizaciones normativas que expliciten los usos permitidos de la IA en evaluación, los mecanismos de auditoría de modelos y los derechos de estudiantes y familias respecto al tratamiento de sus datos.

Implicaciones para la política educativa y la práctica pedagógica en Ambato

En el contexto de Ambato y la provincia de Tungurahua, donde convergen instituciones con diversas condiciones de infraestructura y acceso a tecnologías, la introducción de sistemas de machine learning para evaluación formativa podría integrarse gradualmente en el marco de la estrategia nacional de transformación digital del sistema educativo. Los resultados de este estudio sugieren que comenzar por pilotos acotados en instituciones con cierta base tecnológica y con profesorado voluntario permitiría ajustar modelos, instrumentos y flujos de trabajo, antes de considerar escalas mayores. Además, articular estos pilotos con el ecosistema de datos abiertos del Ministerio de Educación facilitaría construir modelos que, sin vulnerar la privacidad, se beneficien de información contextual más amplia para contextualizar los resultados de aula.

En términos de práctica pedagógica, el uso de reportes automatizados que priorizan estudiantes en riesgo y destrezas con mayores dificultades puede apoyar la planificación de actividades de refuerzo diferenciadas, grupos de apoyo focalizados y ajustes curriculares en tiempo real, fortaleciendo el carácter adaptativo de la enseñanza. La literatura sobre IA en evaluación formativa coincide en que el mayor valor de estas herramientas está en la capacidad de ofrecer ciclos más cortos de retroalimentación, lo que favorece la autorregulación del aprendizaje y la participación activa del estudiantado en el seguimiento de su progreso. En el subnivel de Básica Elemental, esto podría traducirse en estrategias simples, como mostrar a las niñas y niños representaciones visuales de su avance en determinados criterios, siempre mediadas por la explicación y contención del docente para evitar comparaciones estigmatizantes.

Al mismo tiempo, los hallazgos sugieren que la formación docente en Ambato y en el conjunto del país debería incorporar de manera más explícita competencias en analítica de aprendizaje, ética de la IA y uso pedagógico de datos, complementando los cursos actuales sobre transformación digital y diseño de actividades en entornos virtuales. Esta formación no solo debería centrarse en el manejo técnico de herramientas, sino en la capacidad crítica para interpretar resultados, identificar posibles sesgos y tomar decisiones pedagógicas fundamentadas. El fortalecimiento de estas capacidades es clave para que la IA no sea

percibida como una “caja negra” que impone diagnósticos, sino como un recurso cuya lógica el profesorado comprende y gobierna.

Limitaciones del estudio y líneas futuras de investigación

Entre las principales limitaciones se encuentra el carácter piloto de la muestra docente y estudiantil, así como la concentración en instituciones con cierto nivel de digitalización, lo que restringe la generalización de los resultados a contextos con menor infraestructura tecnológica. Del mismo modo, el uso de registros institucionales de desempeño puede estar afectado por variaciones en criterios de calificación, prácticas de registro y calidad de los datos, lo que introduce ruido en el entrenamiento de los modelos de machine learning y puede afectar su desempeño cuando se transfieren a otras instituciones. Una línea de investigación posterior debería explorar métodos de normalización y calibración que permitan armonizar datos de distintas fuentes, así como analizar la sensibilidad de los modelos a estas variaciones.

Otra limitación es que el estudio se concentra en medidas de rendimiento académico cuantificables y no incorpora de forma sistemática indicadores socioemocionales, de participación familiar u otras dimensiones que, según la literatura, influyen de manera importante en el aprendizaje en Educación Básica. Integrar datos más ricos y multidimensionales podría mejorar la capacidad de los modelos para identificar situaciones de riesgo menos visibles en las calificaciones, aunque también plantea desafíos mayores en términos de privacidad y consentimiento informado. Asimismo, sería pertinente realizar estudios longitudinales que examinen el impacto sostenido de la utilización de modelos de machine learning en la práctica docente, el rendimiento estudiantil y la cultura evaluativa institucional, más allá de los efectos inmediatos sobre la carga de trabajo.

Finalmente, futuras investigaciones podrían profundizar en el diálogo entre los marcos normativos nacionales de evaluación estudiantil, las orientaciones ministeriales sobre IA y las experiencias docentes en contextos específicos como Ambato, buscando construir lineamientos de buenas prácticas que hagan compatible la innovación tecnológica con los principios de equidad, inclusión y buen vivir que orientan el sistema educativo ecuatoriano. Este enfoque permitiría avanzar hacia un modelo de evaluación formativa apoyada en IA que no solo sea técnicamente eficiente, sino también socialmente aceptable y pedagógicamente transformador.

5. CONCLUSIONES

La investigación muestra que la incorporación de algoritmos de machine learning en la evaluación formativa de estudiantes de Educación General Básica, subnivel Básica Elemental, en instituciones de Ambato, constituye una alternativa viable para disminuir la carga operativa docente sin desnaturalizar el sentido pedagógico de la evaluación. Los resultados indican que es posible reducir entre 1,5 y 2,5 horas semanales de trabajo asociado principalmente a la corrección y elaboración de informes, al automatizar la agregación de datos, la generación de reportes y la identificación de estudiantes en riesgo, respetando el marco normativo que otorga a la evaluación formativa un peso central en la calificación.

En términos de rendimiento técnico, los modelos de machine learning implementados,

especialmente el bosque aleatorio, alcanzaron niveles de exactitud y capacidad discriminativa comparables con estudios recientes sobre predicción de resultados académicos y sistemas de alerta temprana, lo que respalda su pertinencia para apoyar decisiones pedagógicas en contextos reales de aula. La relevancia de variables como el porcentaje de tareas entregadas, el desempeño en actividades formativas intermedias y los resultados de diagnósticos iniciales refuerza la idea de que el seguimiento sistemático del trabajo cotidiano del estudiantado constituye un insumo clave para anticipar dificultades y diseñar intervenciones oportunas. De este modo, los modelos se alinean con el carácter continuo y procesual de la evaluación formativa promovido por la política educativa ecuatoriana.

Desde la perspectiva docente, la IA es percibida como un apoyo potencialmente valioso para organizar la información evaluativa, priorizar casos y mejorar la retroalimentación, siempre que se mantenga la centralidad del juicio profesional y se establezcan salvaguardas éticas claras. Las y los docentes enfatizan que los resultados algorítmicos deben entenderse como señales que orientan la atención y no como etiquetas deterministas, insistiendo en la necesidad de evitar usos punitivos o estigmatizantes de las predicciones, en sintonía con las orientaciones nacionales sobre uso pedagógico y ético de herramientas de IA. Este posicionamiento coincide con marcos internacionales que proponen concebir la IA como un componente de modelos híbridos de evaluación, donde la automatización se combina con la interpretación humanizada de los datos.

Para la política educativa, los hallazgos sugieren que la integración de IA en la evaluación formativa puede articularse con la agenda de transformación digital del sistema nacional, comenzando por pilotos en instituciones con condiciones mínimas de infraestructura y capacidad de gestión de datos. La articulación entre los desarrollos algorítmicos y el ecosistema de datos abiertos del Ministerio de Educación, así como con iniciativas de formación docente en evaluación e IA, podría consolidar un marco de uso responsable y efectivo de estas tecnologías, contribuyendo a una evaluación más oportuna, criterial y centrada en el aprendizaje. No obstante, se requiere avanzar en lineamientos específicos que regulen el uso de IA en evaluación, definan estándares de transparencia, rendición de cuentas y protección de datos, y garanticen que las decisiones finales sigan siendo pedagógicamente fundamentadas.

En cuanto a las contribuciones al campo académico, el estudio aporta evidencia empírica situada en un contexto urbano interandino ecuatoriano, complementando revisiones y experiencias previas que habían abordado la evaluación formativa con IA principalmente en educación media y superior o en otros países de la región. Al focalizarse en Básica Elemental, se pone de relieve que la IA puede apoyar procesos evaluativos desde los primeros años de escolaridad, siempre que los modelos se diseñen con criterios de interpretabilidad, gradualidad en la implementación y adecuación a las particularidades del desarrollo infantil. Esta mirada contribuye a desplazar el énfasis exclusivamente tecnologicista y subraya la importancia de marcos didácticos y socioemocionales que orienten el uso de la IA en la infancia.

Entre las líneas de investigación futuras se vislumbran al menos cuatro direcciones. Primero, ampliar la escala de los estudios a muestras más grandes y diversas de instituciones de Ambato y de otras provincias, explorando la estabilidad de los modelos cuando se enfrentan

a currículos, culturas escolares e infraestructuras distintas. Segundo, incorporar variables socioemocionales, de participación familiar y de contexto comunitario en los modelos, con el fin de capturar dimensiones no estrictamente académicas pero decisivas para el aprendizaje, cuidando las implicaciones éticas de manejar información sensible. Tercero, realizar estudios longitudinales que analicen el impacto sostenido de la utilización de IA en la cultura evaluativa institucional, la motivación estudiantil y la agencia docente, más allá del ahorro de tiempo inmediato. Finalmente, profundizar en la construcción de marcos normativos y guías de buenas prácticas que integren la visión del Ministerio de Educación, de las instituciones formadoras de docentes y de las comunidades escolares, para consolidar un modelo de evaluación formativa apoyada por IA que sea técnicamente sólido, éticamente responsable y socialmente contextualizado.

6. REFERENCIAS BIBLIOGRÁFICAS

Aimara, M. S. C., Campos, N. G. C., Paredes, E. A. G., & Gordon, B. I. N. (2025). Inteligencia artificial en la evaluación formativa de la educación básica: estrategias para mejorar el progreso académico de los estudiantes. Polo del Conocimiento, 10(4), 680-697

Alkan, B. B., Kuzucuk, S., Alkan, N., & Sinan, A. (2025). Using machine learning to predict student outcomes for early intervention and formative assessment. Scientific Reports, 15, 39797. https://doi.org/10.1038/s41598-025-23409-w

Aplicación de la inteligencia artificial en la evaluación. (2025). Innovarium, 1(1). Revinde. https://revinde.org/index.php/innovarium/article/view/31

Baldeón Medina, P. J., González Yagual, T. B., Izquierdo Corozo, C. D., & Villalba Vélez, K. M. (2025). Aplicación de IA en Educación Básica para fortalecer el aprendizaje y la motivación escolar. Horizonte Cientifico International Journal, 3(2), 1-9. https://doi.org/10.64747/dbkg2w76

Frontiers in Education. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. Frontiers in Education, 8, 1270700. https://doi.org/10.3389/feduc.2023.1270700

Fuentes Cabrera, C. J., Tapia Zurita, O. G., & Tapia Herrera, D. V. (2025). Uso de la inteligencia artificial en la evaluación formativa y su incidencia en el desempeño académico de los estudiantes. Revista Social Fronteriza, 5(1), e-611. https://doi.org/10.59814/resofro.2025.5(1)611

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. Computers and Education: Artificial Intelligence, 4, 100130. https://doi.org/10.1016/j.caeai.2023.100130

Guamán Tonato, E. G., Armas Cajas, M. de las M., Aimacaña Crespata, E. P., & Aimacaña Crespata, M. O. (2025). La ética como eje del uso de la inteligencia artificial en los procesos educativos del Ecuador contemporáneo. Horizonte Científico International Journal, 3(2), 1-19. https://doi.org/10.64747/zp5wat16

Innovación educativa – Documentos. (2024). Ministerio de Educación, Deporte y Cultura del Ecuador. https://educacion.gob.ec/documentos-innovacion-educativa/

Instituto Nacional de Evaluación Educativa (Ineval). (2025). Evaluación nacional Ser Estudiante 2024: Resultados preliminares. https://educacion.gob.ec/ineval-presentara-los-resultados-de-la-evaluacion-nacional-ser-estudiante-2024/

La transformación digital en el Sistema Nacional de Educación. (2025). Ministerio de Educación, Deporte y Cultura del Ecuador. https://educacion.gob.ec/transformacion-digital-en-el-sistema-nacional-de-educacion/

La transformación educativa prioriza el aprendizaje. (2024). Ministerio de Educación, Deporte y Cultura del Ecuador. https://educacion.gob.ec/la-transformacion-educativa-prioriza-el-aprendizaje/
Lineamientos para la evaluación en formación técnica. (2022). Ministerio de Educación, Deporte y Cultura del Ecuador. https://recursos.educacion.gob.ec/wp-content/uploads/2022/curriculares/BT1_Lineamientos_Evaluacion_Formacion_Tecnica.pdf
Ministerio de Educación. (2015). Estadística educativa: Volumen 1 (PUB_EstadisticaEducativaVol1). https://educacion.gob.ec/wp-content/uploads/downloads/2017/06/PUB_EstadisticaEducativaVol1_mar2015.pdf
Ministerio de Educación. (2017). Estadística educativa: Volumen 2. https://educacion.gob.ec/wp-content/uploads/downloads/2021/10/PUB_EstadisticaEducativa_Vol2.pdf
Ministerio de Educación. (2022). Estadística educativa: Volumen 3. https://educacion.gob.ec/wp-content/uploads/downloads/2022/12/Estadistica-Educativa-Vol3.pdf
Ministerio de Educación. (2023). Estadística educativa: Volumen 4. https://educacion.gob.ec/wp-content/uploads/downloads/2023/11/Estadistica-Educativa_Volumen-4.pdf
Ministerio de Educación. (2024). Instructivo de evaluación estudiantil Costa-Galápagos 2024. https://educacion.gob.ec/wp-content/uploads/downloads/2024/08/instructivo-evaluacion-estudiantil-costa-galapagos-2024.pdf
Ministerio de Educación. (2024). Instructivo de evaluación estudiantil Sierra-Amazonía 2024–2025. https://educacion.gob.ec/wp-content/uploads/downloads/2024/10/instructivo-evaluacion-estudiantil-Sierra-Amazonia-2024-2025.pdf
Ministerio de Educación. (2025). Datos abiertos de educación. Ministerio de Educación, Deporte y Cultura del Ecuador. https://educacion.gob.ec/datos-abiertos/
Ministerio de Educación. (2025). Datos abiertos – Recursos educativos. Portal de Recursos Educativos Digitales. https://recursos.educacion.gob.ec/datos-abiertos/
Ministerio de Educación. (2025). Instructivo de evaluación estudiantil 2025. https://educacion.gob.ec/wp-content/uploads/downloads/2025/04/Instructivo-de-Evaluacion-Estudiantil-2025.pdf
Ministerio de Educación. (s. f.). Estrategia de nivelación formativa. Portal de Recursos Educativos Digitales. https://recursos.educacion.gob.ec/red/estrategia-de-nivelacion-formativa/
Ministerio de Educación. (s. f.). Evaluación formativa: Evaluando continuamente a los estudiantes. Plataforma MeCapacito. https://mecapacito.educacion.gob.ec/evaluacion-formativa/
Ministerio de Educación. (s. f.). Orientaciones para el uso pedagógico de herramientas de Inteligencia Artificial en el proceso de enseñanza-aprendizaje. https://recursos.educacion.gob.ec/wp-content/uploads/REDA/AED/orientaciones_uso_pedagogico_ia.pdf
Ministerio de Educación. (s. f.). Portal de recursos educativos digitales. https://recursos.educacion.gob.ec

Originalidad: Los autores declaran que el trabajo es inédito y no ha sido publicado, ni parcial ni totalmente, en otra publicación.

Conflicto de intereses: Los autores declaran que no existen conflictos de intereses.
Aprobación ética e institucional: El estudio no contó con aprobación de un comité de ética ni con autorización institucional. Los autores declaran que la participación se realizó mediante consentimiento informado y que los datos analizados fueron anonimizados.
Financiamiento: La investigación no recibió financiamiento externo.
Disponibilidad de datos: Las bases anonimizadas, los instrumentos y los consentimientos se encuentran bajo custodia de los autores y están disponibles mediante solicitud razonable al autor de correspondencia, con sujeción a las obligaciones de confidencialidad aplicables.
Uso de inteligencia artificial: Los autores declaran que no utilizaron herramientas de inteligencia artificial generativa en la elaboración del manuscrito.
Contribuciones de autoría (CRediT): Kevin David Cruz-Nicolalde realizó la conceptualización, metodología, investigación, análisis formal, curación de datos, redacción del borrador y revisión y edición del manuscrito.
Historial editorial: recibido el 6 de julio de 2026; revisado el 30 de julio de 2026; aceptado el 21 de agosto de 2026.