



APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA EVALUACIÓN DE ESTUDIOS AMBIENTALES

Evelyn Rosalía Meres Morales¹
ORCID: 0000-0001-9016-2721

¹ Doctoranda en Informática, Universidad Americana de Europa (UNAE). Correo: ermeresm@gmail.com
[Revista de Investigación Multidisciplinaria Iberoamericana, RIMI](#) © 2023 by Elizabeth Sánchez Vázquez is licensed under

RESUMEN

La evaluación de estudios ambientales en el Perú fue una función crucial para proyectos de inversión públicos y privados, pues el resultado final determinó si el proyecto se ejecutó o no. Este proceso fue engorroso y tomó entre 3 y 12 meses, según la envergadura y el ámbito del proyecto. El principal problema radicó en el gran tamaño de los expedientes, que debieron revisarse manualmente y, además, pudieron requerir actualizaciones, lo que obligó a reiniciar la revisión. En el sector minero, la evaluación fue crítica para asegurar el desarrollo sostenible y el cumplimiento normativo; por ello, la investigación se justificó por la necesidad de reducir tiempos y costos, minimizar errores humanos, asegurar el cumplimiento normativo, aportar científica y tecnológicamente, y promover una reflexión ética y técnica.

El objetivo fue diseñar e implementar un modelo híbrido de inteligencia artificial que combinó conocimiento experto ambiental con algoritmos de aprendizaje automático, sustentado en datos históricos y sintéticos generados con apoyo de ChatGPT y sometidos a controles de coherencia técnica, lógica y normativa, para predecir y explicar los resultados de evaluaciones ambientales en minería. La metodología se estructuró en cuatro fases: marco conceptual y técnico, preparación de datos, desarrollo/validación/evaluación de modelos, y validación cualitativa con expertos.

Como resultados destacados, Random Forest mostró exactitud de 96.9% y alta estabilidad en la segunda iteración, fue eficaz ante variables heterogéneas y ruido, y destacó por su interpretabilidad, clave para el sector regulatorio. En conclusión, la IA presenta potencial para mejorar, y Random Forest se perfiló como el modelo más adecuado por su interpretabilidad, estabilidad y potencial para reducir errores humanos.

Palabras clave: Estudios Ambientales, Inteligencia Artificial, Random Forest

ABSTRACT

Environmental assessment in Peru was a crucial function for public and private investment projects, as the final outcome determined whether a project was carried out or not. This process was cumbersome and took between 3 and 12 months, depending on the scale and scope of the project. The main problem lay in the large size of the files, which had to be reviewed manually and could also require updates, forcing the review process to restart. In the mining sector, the assessment was critical to ensure sustainable development and regulatory compliance; therefore, the research was justified by the need to reduce time and costs, minimize human error, ensure regulatory compliance, contribute scientifically and technologically, and promote ethical and technical reflection.

The objective was to design and implement a hybrid artificial intelligence model that combined expert environmental knowledge with machine learning algorithms, based on historical and synthetic data generated with the support of ChatGPT and subjected to technical, logical and regulatory consistency controls, in order to predict and explain the outcomes of environmental assessments in mining. The methodology was structured into four phases: conceptual and technical framework, data preparation, model development/validation/evaluation, and qualitative validation with experts.

As for the main results, Random Forest showed an accuracy of 96.9% and high stability in the second iteration, proved effective in handling heterogeneous variables and noise, and stood out for its interpretability, which is key in the regulatory sector. In conclusion, AI has the potential to improve, and Random Forest emerged as the most suitable model due to its interpretability, stability, and potential to reduce human error.

Keywords: Environmental Studies, Artificial Intelligence, Random Forest

INTRODUCCIÓN

La evaluación de estudios ambientales es una función crucial para los proyectos de inversión públicos y privados, el resultado final que se libere determinará si el proyecto se llevará a cabo o no. Se examinan estudios ambientales de distintos sectores tales como: transportes, salud, construcción, agricultura, saneamiento, electricidad, hidrocarburos y minería, cada uno de ellos emplea distintos criterios para su valoración. El presente estudio se delimita al sector minero peruano.

Es un proceso engorroso que puede tomar entre 3 y 12 meses, dependiendo de la envergadura y ámbito del proyecto de inversión. El principal reto radica en el gran tamaño de los expedientes, los cuales deben ser revisados manualmente y, además, pueden requerir actualizaciones, lo que obliga a reiniciar el proceso de revisión. En esta revisión participan distintos especialistas quienes proveen su retroalimentación. En caso de existir observaciones, éstas se deben subsanar dentro de un plazo específico de acuerdo a las Normativas.

Este proceso no solo retrasa la aprobación de proyectos de inversión críticos para el desarrollo económico, sino que también puede generar inconsistencias en la revisión debido al volumen de información y la intervención manual de distintos especialistas.

El presente estudio no solo propone automatizar la evaluación de estudios ambientales mediante inteligencia artificial, sino también desarrollar un enfoque metodológico replicable, que contribuya al campo de la IA mediante la integración de modelos híbridos, generación de datos sintéticos ambientalmente coherentes (producidos a partir del manual de evaluación de estudios ambientales y los resultados de evaluaciones históricas, lográndose estructurar la información con datos generales del estudio, impactos ambientales, medidas de mitigación, observaciones, indicadores ambientales, participación ciudadana, y obligaciones ambientales) y una arquitectura multi-modelo que permita recomendaciones automatizadas y explicables.

La evaluación de estudios ambientales (EIA), especialmente en el sector minero, es un proceso crítico para garantizar el desarrollo sostenible y el cumplimiento normativo de los proyectos de inversión. Sin embargo, este proceso suele ser complejo, prolongado y dependiente de la revisión manual de grandes volúmenes de información técnica, lo que genera demoras, inconsistencias y un elevado uso de recursos humanos y económicos.

En este contexto, la inteligencia artificial (IA) representa una alternativa innovadora para transformar la forma en que se evalúan los estudios ambientales. Esta investigación propone el diseño de un modelo híbrido que combine el conocimiento experto basado en normativa ambiental con algoritmos de aprendizaje automático, complementado con un generador de datos sintéticos ambientalmente coherentes y una arquitectura multi-modelo que permita no solo predecir la aprobación o no conformidad de un expediente, sino también recomendar acciones correctivas específicas.

La naturaleza de este estudio es tanto exploratoria como aplicada. Exploratoria, porque busca adaptar y evaluar herramientas de IA en un dominio regulado con alta complejidad normativa; aplicada, porque tiene como objetivo implementar una solución tecnológica real que mejore la eficiencia, consistencia y capacidad explicativa del proceso de evaluación ambiental. Se utilizarán técnicas de machine learning, para automatizar la revisión de criterios técnicos, analizar patrones históricos y generar retroalimentación inmediata.

Los principales motivos que justifican esta investigación son: reducción de tiempos y costos, minimización de errores humanos, aseguramiento del cumplimiento normativo, contribución científica y tecnológica, reflexión ética y técnica.

El objetivo de la presente investigación es diseñar e implementar un modelo híbrido de inteligencia artificial que combine conocimiento experto ambiental con algoritmos de aprendizaje automático, sustentado en datos históricos y sintéticos validados, para predecir y explicar el resultado de las evaluaciones de estudios ambientales del sector minero. El modelo buscará identificar patrones, aplicar criterios normativos de forma automatizada y reducir significativamente los tiempos de revisión, aportando valor tanto al campo de la inteligencia artificial como al desarrollo eficiente y sostenible de proyectos sujetos a Declaración de Impacto Ambiental.

Los Objetivos específicos son:

- ☐ Diseñar un modelo híbrido que combine conocimiento experto ambiental con algoritmos de aprendizaje automático para evaluar proyectos mineros.
- ☐ Crear un generador de datos sintéticos que simule expedientes con realismo técnico y normativo.
- ☐ Implementar una arquitectura multi-modelo que integre predicción de resultados y recomendaciones automáticas.
- ☐ Comparar el rendimiento de modelos de IA (Random Forest, Gradient Boosting) con métodos tradicionales y permitir aprendizaje continuo para mejorar desempeño.
- ☐ Evaluar el impacto del modelo en la reducción de tiempos de evaluación y analizar posibles sesgos, proponiendo estrategias de mitigación.

¿Cómo puede diseñarse un modelo de inteligencia artificial que, integrando conocimiento experto, datos sintéticos y una arquitectura multi-modelo, mejore la exactitud, la interpretabilidad y la eficiencia del proceso de evaluación de estudios ambientales en el sector minero, aportando al mismo tiempo una metodología innovadora al campo de la inteligencia artificial?

Si se desarrolla un modelo híbrido que integre reglas expertas ambientales con algoritmos de aprendizaje automático, complementado con datos sintéticos ambientalmente coherentes y una arquitectura que permita predecir resultados y generar recomendaciones, entonces se mejorará la capacidad predictiva y explicativa de los sistemas de evaluación ambiental automatizada, contribuyendo con un nuevo enfoque metodológico a la inteligencia artificial.

ESTADO DEL ARTE

La inteligencia artificial (IA) ha mostrado un gran potencial para mejorar la Evaluación de Impacto Ambiental (EIA), facilitando el análisis de grandes volúmenes de datos y reduciendo errores humanos en distintos sectores, especialmente en ingeniería civil y minería (Zambrano Salazar, 2024). Técnicas como Random Forest, Gradient Boosting, AHP, SVM y AutoML permiten predecir resultados, detectar patrones complejos y generar evaluaciones más rápidas y precisas que los métodos tradicionales.

Además, la IA contribuye a minimizar sesgos humanos y permite procesar información no estructurada mediante Procesamiento de Lenguaje Natural (Monteagudo, 2023; Liu, 2023). Estudios internacionales destacan aplicaciones prácticas: sistemas en Canadá procesan miles de EIA con tiempos significativamente menores, y proyectos en la UE y la OECD desarrollan estándares para IA explicable y transparente en evaluación ambiental (Tremblay, 2023; European Environment Agency, 2023; OECD, 2023).²

No obstante, se identifican limitaciones: necesidad de datos de alta calidad, complejidad computacional, dependencia tecnológica y adaptación continua ante cambios en las condiciones ambientales. La investigación actual confirma que, con estrategias adecuadas de validación y modelado, la

² Agencia europea de medio ambiente

[Revista de Investigación Multidisciplinaria Iberoamericana. RIMI © 2023 by Elizabeth Sánchez Vázquez is licensed under](#)

IA puede transformar la evaluación ambiental, incrementando exactitud, eficiencia y confiabilidad en la toma de decisiones.

MARCO TEÓRICO

En las últimas décadas, la Inteligencia Artificial (IA) emerge como una de las disciplinas más innovadoras y dinámicas dentro del campo de la informática, impactando diversos sectores como la salud, la educación, la industria y los servicios. Su capacidad para procesar grandes volúmenes de datos, aprender de ellos y tomar decisiones basadas en patrones complejos revoluciona la forma en que interactuamos con la tecnología y cómo ésta influye en nuestras vidas cotidianas.

Esta investigación se enmarca en el campo de estudio de la inteligencia artificial (IA) y sus aplicaciones prácticas, con énfasis en fundamentos como el aprendizaje automático y la estadística. Este campo no solo se centra en el desarrollo y perfeccionamiento de algoritmos y modelos matemáticos que subyacen a la IA, sino también en la exploración de sus múltiples aplicaciones en entornos reales.

Inteligencia Artificial (IA)

La *Inteligencia Artificial (IA)* es el área de la computación que desarrolla sistemas capaces de realizar tareas que requieren capacidades cognitivas asociadas con la inteligencia humana, como razonar, aprender, interpretar datos complejos y tomar decisiones basadas en información. Estas capacidades se logran mediante algoritmos que aprenden de los datos y se adaptan con el tiempo para mejorar su desempeño en tareas específicas. (Kumar, 2024)

Machine Learning (ML)

El *Machine Learning (ML)* es un subcampo de la IA que se centra en desarrollar algoritmos capaces de aprender automáticamente a partir de datos sin ser programados explícitamente para cada tarea. Los modelos de ML identifican patrones en los datos, extraen conocimiento útil y mejoran su desempeño a medida que reciben más información, permitiendo automatizar tareas como clasificación, predicción y optimización de decisiones. (Mosqueira-Rey, 2023)

Procesamiento de Lenguaje Natural (PLN / NLP)

El *Procesamiento de Lenguaje Natural (PLN)* es una rama de la IA que combina métodos de lingüística, aprendizaje automático y modelos computacionales para que los sistemas interpreten, procesen y generen lenguaje humano, tanto en texto como en voz. El PLN se aplica en traducción automática, reconocimiento de voz, análisis de sentimientos y generación de texto, facilitando la interacción natural entre humanos y máquinas (Zubiaga, 2024)

Evaluación de Impacto Ambiental (EIA)

A partir de mediados de siglo pasado, ciudadanos, empresas y gobiernos empiezan a diseñar los proyectos sobre la base de modelos de gestión sistemáticos que ayudan a prevenir daños al ambiente. En la década de 1960, en Europa y en los Estados Unidos, llamados de atención acerca de los efectos de las actividades humanas sobre el ambiente estimulan el interés por estudiar los impactos ambientales y buscar formas de prevenirlos.

METODOLOGÍA

Enfoque general

La presente investigación adoptó un enfoque aplicado y mixto, orientado a resolver un problema concreto: la lentitud, complejidad y variabilidad en la evaluación de estudios de impacto ambiental (EIA) para proyectos de inversión en el Perú. Se buscó diseñar, desarrollar y validar modelos de inteligencia artificial (IA), basados en aprendizaje automático (Machine Learning) y procesamiento de lenguaje natural (PLN), que apoyaran y optimizaran este proceso.

El diseño mixto secuencial explicativo (QUAN → qual) combinó el análisis cuantitativo (medición del rendimiento de los modelos de IA) con análisis cualitativo (percepción de expertos sobre la utilidad y aplicabilidad de estos modelos), permitiendo una comprensión integral del problema y de la solución propuesta.

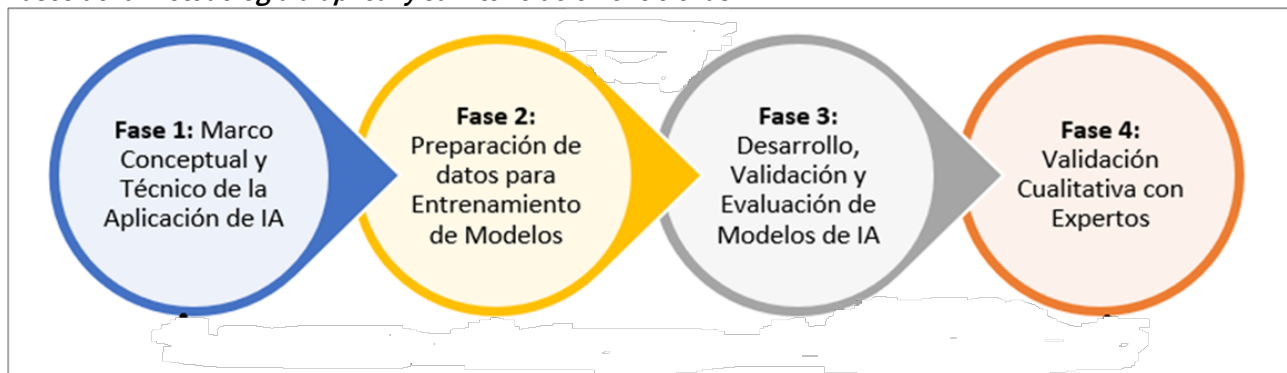
La integración del análisis cuantitativo y cualitativo permitió no solo desarrollar modelos predictivos, sino también entender cómo estos modelos fueron percibidos y podrían ser interpretados o mejorados con base en la experiencia de los expertos.

Diseño

La presente investigación tuvo como propósito resolver problemas reales en la evaluación de estudios ambientales, reduciendo el tiempo de evaluación y mejorando la exactitud, estuvo basada en datos históricos de estudios ambientales y en la experimentación con algoritmos de IA. Se aplicó una metodología mixta que incluyó fases iterativas para ajustar los modelos a través de experimentación constante, refinando los algoritmos a medida que se probaron y optimizaron; por lo tanto, el tipo de investigación a realizar fue Aplicada.

Figure 1

Fases de la Metodología a aplicar y su interrelación entre ellas



Nota: Elaboración propia, 2026

La Figura 1 presenta una síntesis de las fases de la metodología a aplicar.

Fases de la metodología

Fase 1. Marco conceptual y Técnico de la Aplicación de IA

En esta primera fase, el objetivo fue establecer las bases conceptuales y metodológicas del proyecto, identificando el alcance, los objetivos y el contexto técnico-científico relacionado con la aplicación de IA en la evaluación de impactos ambientales.

Fase 2. Generación de datos/preparación del dataset para entrenamiento de modelos

El conocimiento experto ambiental fue incorporado al modelo computacional durante la etapa de construcción del conjunto de datos sintéticos. Para ello, se utilizaron los criterios técnicos, variables de evaluación y consideraciones metodológicas establecidas en el Manual de Evaluación de Estudios Ambientales del sector minero como fuente de conocimiento experto para definir la estructura de los registros generados. Estos criterios permitieron establecer las variables predictoras relacionadas con las características del proyecto, impactos ambientales, medidas de manejo, cumplimiento normativo y condiciones técnicas del expediente.

A partir de este conocimiento, se definieron escenarios sintéticos representativos de expedientes ambientales y sus respectivas categorías de salida (Aprobado/No conforme), buscando mantener coherencia con los criterios utilizados en la evaluación ambiental. De esta manera, el conocimiento experto no fue incorporado como una modificación manual del algoritmo, sino como parte del diseño del dataset de entrenamiento, permitiendo que los modelos de aprendizaje automático aprendieran patrones asociados a criterios ambientales previamente establecidos.

Esta fase entregó el conjunto de datos listo para el entrenamiento, validación y prueba de los modelos IA en la siguiente fase.

Fase 3. Desarrollo, validación y evaluación de modelos de IA

En esta fase, se desarrollaron y evaluaron dos modelos de IA (Random Forest y Gradient Boosting) para determinar cuál fue más adecuado para los objetivos planteados. Se incluyó un análisis detallado de su rendimiento y capacidad de interpretación.

Fase 4: Validación cualitativa con expertos

Esta fase tuvo como objetivo evaluar la interpretabilidad, aceptabilidad y aplicabilidad práctica de los modelos IA por parte de evaluadores humanos.

En esta fase participaron 2 especialistas en evaluación ambiental del sector minero, seleccionados mediante muestreo intencional por criterios de experiencia (5 años mínimos, experiencia en revisión de DIA/EIA, conocimiento normativo). Se aplicó una ficha de validación experta con escala Likert de cinco puntos, dirigida a especialistas con experiencia en evaluación ambiental. El instrumento evaluó cuatro dimensiones: interpretabilidad del modelo, coherencia técnica y normativa, utilidad práctica y confianza en su aplicabilidad. Cada dimensión incluyó ítems cerrados valorados del 1 al 5, donde 1 representó “totalmente en desacuerdo” y 5 “totalmente de acuerdo”. Adicionalmente, se incorporó un campo de observaciones abiertas para recoger recomendaciones, limitaciones percibidas y sugerencias de mejora.

El análisis de la validación cualitativa combinó estadística descriptiva y análisis temático simple. Las respuestas en escala Likert fueron procesadas mediante promedios por dimensión, considerando interpretabilidad, utilidad, confiabilidad, aplicabilidad y riesgos. Las respuestas abiertas fueron agrupadas por categorías para identificar fortalezas, limitaciones y recomendaciones de mejora. Finalmente, estos hallazgos se contrastaron con las métricas cuantitativas del modelo, con el fin de valorar su viabilidad como herramienta de apoyo a la evaluación ambiental.

Población y muestra

La población estuvo conformada por 3,754 proyectos de inversión evaluados entre el 2016 y el 2024. Como muestra se tomó el sector Minería por tener la mayor cantidad de IGAs y ser el sector con mayor repercusión económica en los proyectos de inversión en el país.

De los 1,642 proyectos en Minería se tomaron 312 proyectos de inversión evaluados, considerando un índice de confianza del 95% y margen de error del 5%.

De 312 proyectos solo se encontraron 6 de tipo EIAd, todos aprobados, lo cual no es suficiente para el modelo, por lo tanto, se generaron datos sintéticos balanceados.

Datos sintéticos generados con apoyo de Chatgpt y sometidos a controles de coherencia técnica, lógica y normativa:

Los datos sintéticos fueron generados con apoyo de ChatGPT, utilizando instrucciones estructuradas basadas en el Manual para la Evaluación de Estudios Ambientales del Subsector Minería, expedientes disponibles en el Sistema de Evaluación Ambiental en Línea SEAL y la normativa ambiental minera vigente. Previamente, se definieron las variables, categorías, rangos y reglas de relación correspondientes a datos generales del proyecto, impactos ambientales, medidas de mitigación, observaciones, cumplimiento normativo, indicadores ambientales y participación ciudadana.

A partir de estas especificaciones, ChatGPT fue utilizado como herramienta para generar 920 casos sintéticos. Su función se limitó a producir registros conforme a las reglas proporcionadas y no a establecer criterios normativos ni adoptar decisiones ambientales. Posteriormente, la base fue revisada para identificar registros duplicados, valores fuera de rango, categorías inconsistentes y relaciones no plausibles entre impactos, medidas de mitigación, observaciones y resultado de la evaluación.

Debido a que los datos fueron generados artificialmente, estos no representan necesariamente la distribución real de los expedientes mineros. Por ello, los resultados obtenidos se interpretaron como evidencia experimental y deberán contrastarse progresivamente con una mayor cantidad de expedientes reales y con la revisión de especialistas ambientales.

Tabla 1

Variables y sus fuentes normativas

Variable del dataset	Fuente	Elemento extraído
Factor, nivel y ubicación	Manual EIA-d Minería	Evaluación de impactos
Cumple / No cumple	D. S. N.° 040-2014-EM	Obligaciones ambientales
Convocatoria y mecanismos	D. S. N.° 028-2008-EM	Participación ciudadana
Talleres informativos	R. M. N.° 304-2008-MEM/DM	Talleres y audiencias
Línea base, área de influencia, impactos	R. M. N.° 116-2015-MEM/DM	Contenido mínimo

Tratamiento de la distribución de clases

El conjunto de datos sintéticos estuvo conformado por 920 expedientes, distribuidos equitativamente entre 460 proyectos aprobados y 460 proyectos no conformes. El balance se obtuvo mediante la incorporación de 420 expedientes sintéticos no conformes, generados con apoyo de ChatGPT a partir de criterios técnicos y normativos del sector minero, que complementaron los 40 casos no conformes

inicialmente disponibles. Al incorporar seis expedientes reales aprobados, el conjunto integrado quedó conformado por 466 proyectos aprobados 50.32 % y 460 proyectos no conformes 49.68 %. Debido a que esta diferencia no representó un desbalance significativo, no se aplicó SMOTE en el modelo final.

SMOTE fue considerado inicialmente como una alternativa para corregir la baja representación de los proyectos no conformes. Sin embargo, la incorporación de 420 expedientes sintéticos ambientalmente coherentes permitió equilibrar previamente la variable objetivo. Por ello, aplicar SMOTE sobre el conjunto final habría generado una segunda capa de síntesis innecesaria y podía aumentar el riesgo de sobreestimación del desempeño. En consecuencia, el entrenamiento final se realizó con la distribución original de 466 proyectos aprobados y 460 no conformes, utilizando partición y validación estratificadas.

Tabla 2
Cuadro de variables

Nombre de variable	Definición operacional	Tipo	Codificación	Fuente	Rol
Tipo de minería	Modalidad de explotación minera correspondiente al proyecto evaluado. Categorías encontradas: Subterránea y Tajo abierto.	Categórica nominal	En las pruebas se aplicó LabelEncoder. Para una versión reproducible se recomienda one-hot: tipo_subterránea y tipo_tajo_abierto.	Sintética; real cuando se incorporan expedientes reales	Predictora
Región	Departamento peruano en el que se ubica el proyecto. En el archivo: Arequipa, Ayacucho, Cajamarca, Cusco, Junín y Pasco.	Categórica nominal	One-hot por región. Si se conserva LabelEncoder, debe reportarse el diccionario exacto generado por el código.	Sintética; real cuando se incorporan expedientes reales	Predictora
Fase actual del proyecto	Etapas del ciclo minero en la que se encuentra el proyecto al momento de la evaluación: Exploración, Construcción, Explotación o Cierre.	Categórica nominal	Preferentemente one-hot. En las pruebas originales se reportó LabelEncoder.	Sintética; real cuando se incorporan expedientes reales	Predictora
Monto estimado del proyecto	Valor monetario estimado de la inversión del proyecto, expresado en dólares estadounidenses. En el archivo varía entre aproximadamente USD 5.2 y 99.9 millones.	Numérica continua	Se conserva como valor numérico. Puede aplicarse estandarización o transformación logarítmica si el análisis evidencia asimetría.	Sintética; real cuando se incorporan expedientes reales	Predictora

Nombre de variable	Definición operacional	Tipo	Codificación	Fuente	Rol
Nivel de impacto ambiental	Mayor nivel de impacto identificado entre los impactos registrados para un proyecto. Categorías: Bajo, Medio y Alto.	Categórica ordinal / derivada	Bajo = 1; Medio = 2; Alto = 3. Para consolidar múltiples impactos por proyecto, se propone utilizar el máximo nivel observado .	Derivada de impactos sintéticos o reales	Predictora
Eficacia promedio de las medidas de mitigación	Promedio aritmético del porcentaje de eficacia estimada de todas las medidas de mitigación asociadas a un proyecto. En el archivo, los valores individuales se encuentran entre 70 % y 99 %.	Numérica continua / derivada	Promedio por ID Proyecto: suma de porcentajes de eficacia ÷ número de medidas registradas.	Derivada de medidas sintéticas o reales	Predictora
Cumplimiento normativo	Indicador que representa si el proyecto satisface las obligaciones ambientales registradas.	Binaria / derivada	Regla recomendada: 1 = cumple todas las obligaciones evaluadas; 0 = presenta al menos un incumplimiento . La regla exacta debe coincidir con el código del modelo.	Derivada de registros normativos sintéticos o reales	Predictora
Cantidad de participantes	Número de personas registradas en el mecanismo de participación ciudadana del proyecto. En el archivo varía entre 50 y 549 participantes.	Numérica discreta	Número entero. Si un proyecto tiene varios mecanismos, calcular el promedio o total y declarar expresamente la regla seleccionada.	Sintética; real cuando se incorporan expedientes reales	Predictora
Extensión del impacto	Alcance espacial general del impacto ambiental asociado al proyecto. Categorías: Local, Media y Alta.	Categórica ordinal	Local = 1; Media = 2; Alta = 3.	Sintética; real cuando se incorporan expedientes reales	Predictora
Duración del impacto	Periodo durante el cual se mantiene el impacto ambiental. Categorías: Temporal, Prolongada y Permanente.	Categórica ordinal	Temporal = 1; Prolongada = 2; Permanente = 3.	Sintética; real cuando se incorporan expedientes reales	Predictora

Nombre de variable	Definición operacional	Tipo	Codificación	Fuente	Rol
Reversibilidad del impacto	Posibilidad de recuperar las condiciones ambientales previas después de producido el impacto.	Categórica ordinal	Reversible = 1; Parcialmente reversible = 2; Irreversible = 3.	Sintética; real cuando se incorporan expedientes reales	Predictora
Severidad global	Valoración integrada de la gravedad general de los impactos ambientales del proyecto. Categorías: Baja, Media, Alta y Muy alta.	Categórica ordinal	Baja = 1; Media = 2; Alta = 3; Muy alta = 4.	Sintética; real cuando se incorporan expedientes reales	Predictora
Estado del EIA	Resultado final de la evaluación ambiental del expediente. En el Excel las categorías son Aprobado y No conforme .	Binaria	1 = Aprobado; 0 = No conforme.	Sintética en los 920 registros; real en los expedientes incorporados posteriormente	Objetivo

Las variables predictoras fueron seleccionadas a partir de las dimensiones técnica, económica, ambiental, normativa y social de los expedientes. Los registros provenientes de las distintas tablas fueron consolidados mediante el campo ID Proyecto, que posteriormente fue excluido del entrenamiento. Las variables categóricas fueron transformadas a formato numérico y las variables con múltiples registros por proyecto fueron agregadas mediante reglas previamente definidas. La variable objetivo correspondió al Estado del EIA, codificado como 1 para los proyectos aprobados y 0 para los proyectos no conformes. El cuadro de variables presenta la definición operacional, tipo, codificación, fuente y rol de cada variable.

Arquitectura híbrida y multi-modelo

La solución se estructuró como una arquitectura híbrida modular que integra conocimiento experto ambiental, procesamiento de datos, modelos de aprendizaje automático, técnicas de explicabilidad y supervisión humana. Su carácter híbrido deriva de la combinación entre criterios normativos explícitos y patrones aprendidos automáticamente a partir de los datos. El componente normativo no sustituye al modelo predictivo, sino que interviene en la definición de variables, la generación y verificación de los datos sintéticos, la interpretación de las predicciones y la formulación de observaciones para el evaluador.

La integración del conocimiento experto se realizó en la fase de generación de datos, mediante la incorporación de criterios ambientales provenientes del Manual de Evaluación de Estudios Ambientales. Por tanto, la arquitectura propuesta corresponde a un enfoque de aprendizaje automático supervisado basado en datos construidos bajo criterios técnicos ambientales, donde el conocimiento experto orientó la definición de variables y escenarios de entrenamiento.

La arquitectura comprende siete componentes: entrada y consolidación de datos, conocimiento normativo, preprocesamiento, evaluación de modelos predictivos, selección y configuración

del modelo final, explicabilidad y validación humana. Los datos provenientes de las distintas tablas se integraron mediante el identificador del proyecto y se transformaron en un registro consolidado por expediente. Entre las variables utilizadas se incluyeron tipo de minería, región, fase del proyecto, monto estimado, nivel de impacto, eficacia de las medidas de mitigación, cumplimiento normativo, participación ciudadana, extensión, duración, reversibilidad y severidad global.

En el componente predictivo se implementaron Random Forest y Gradient Boosting como modelos candidatos de clasificación binaria. Ambos recibieron las mismas variables de entrada y generaron como salida una probabilidad y una categoría de resultado: aprobado o no conforme. Los modelos fueron comparados mediante métricas de clasificación y criterios de aplicabilidad, como estabilidad e interpretabilidad. Esta arquitectura no utilizó votación, promedio de probabilidades ni stacking entre ambos algoritmos; el enfoque multi-modelo consistió en su evaluación paralela y posterior selección.

Random Forest fue seleccionado para las etapas posteriores de validación externa, calibración y análisis del umbral de decisión.

Finalmente, las variables influyentes fueron vinculadas con las disposiciones normativas que sustentaron su incorporación al modelo. La predicción y su explicación fueron consideradas insumos para la revisión del especialista ambiental y no una decisión administrativa automática. El sistema contempló el registro de las variables de entrada, versión del modelo, predicción, explicación, referencia normativa y validación humana, con el propósito de asegurar trazabilidad y facilitar auditorías.

RESULTADOS

Table 3

Cuadro comparativo de los resultados obtenidos de los modelos Random Forest y Gradient Boosting

Modelo	Iterac.	Datos	Validación	Exactitud	Precisión por clase	Recall	F1	ROC-AUC	Notas
Random Forest	1	920 sintéticos	Hold-out 70/30	78.26%	—	—	78%	—	Línea base.
Random Forest	2	920 sintéticos + 6 reales	CV estratificada k=5	96.9%	97.1%	96.9%	96.9%	96.9%	Matriz: TN=869, TP=849, FP=37, FN=17.
Gradient Boosting	1	920 sintéticos	Hold-out 80/20	90.09%	90.41%	89.75%	90.08%	96.24%	Línea base sólida.
Gradient Boosting	2	(mismo set GridSearch)	+ —	90.09%	90.41%	89.75%	90.08%	96.24%	Matriz: TN=1667, TP=1647, FP=118, FN=130.

Nota: Elaboración propia, 2026

DISCUSIÓN

El propósito metodológico en resultados corresponde a: Iteración 1 = Línea base, Iteración 2 = Modelo refinado con balanceo y validación cruzada.

De acuerdo al cuadro comparativo Random Forest en Iteración 2 logra Exactitud/F1/ROC-AUC $\approx$ 96.9% con CV estratificada y muy baja tasa de error, Gradient Boosting se queda en $\approx$ 90.09% de Exactitud y AUC $\approx$ 96.24%.

El modelo Random Forest constituye una primera aproximación hacia un sistema automatizado de asistencia a la decisión en el marco de la evaluación ambiental. Su diseño permite integrar criterios normativos, técnicos y participativos, reflejando de forma estructurada la lógica del proceso evaluativo, y ofreciendo una herramienta para priorizar casos, identificar proyectos críticos o validar consistencia entre evaluaciones.

Cabe señalar que los datos son una limitación en la evaluación de estudios ambientales, considerando que en su mayoría son estudios aprobados; por lo tanto, es necesario generar datos sintéticos.

Comparación de Desempeño de los modelos:

Random Forest (RF):

El modelo **Random Forest** mostró un rendimiento sobresaliente en términos de exactitud (96.9%) y estabilidad entre las iteraciones. Este algoritmo es eficaz en entornos donde las variables son heterogéneas y existe un alto grado de ruido, como es común en los estudios ambientales. Además, la interpretabilidad es uno de sus puntos fuertes, lo cual es crucial para el sector regulatorio, ya que permite que los expertos comprendan qué variables son más influyentes en las decisiones de aprobación o rechazo de los proyectos. En términos prácticos, este modelo facilita la automatización de la revisión de los informes ambientales, reduciendo los tiempos de evaluación de proyectos y permitiendo una evaluación más consistente.

Gradient Boosting (GB):

El modelo Gradient Boosting, aunque obtuvo una exactitud similar, tiene la ventaja de ser más preciso en la detección de relaciones no lineales en los datos, lo que puede ser relevante en casos más complejos de evaluación. Sin embargo, como se mencionó, este modelo tiene una ligera desventaja en términos de sensibilidad a los umbrales de decisión. En el contexto ambiental, esto podría implicar que en ciertos casos el modelo clasifique incorrectamente proyectos que no cumplen con los estándares como conformes (falsos positivos) o que no detecte proyectos aptos para su aprobación (falsos negativos). Esto puede ser crucial en la evaluación de proyectos mineros, donde el riesgo de impactos ambientales debe ser cuidadosamente gestionado.

RIESGOS ÉTICOS Y SEGOS POTENCIALES

Pese a los resultados observados, el uso de IA en evaluación ambiental implica riesgos éticos relevantes. En primer lugar, la dependencia de decisiones históricas puede reproducir sesgos previos del sistema evaluador. En segundo lugar, la escasa representación de expedientes desfavorables puede afectar la capacidad del modelo para identificar casos críticos. En tercer lugar, falsos positivos o falsos negativos podrían inducir errores regulatorios con consecuencias ambientales y operativas. Por ello, el modelo debe

[Revista de Investigación Multidisciplinaria Iberoamericana, RIMI](#) © 2023 by Elizabeth Sánchez Vázquez is licensed under

entenderse como herramienta de apoyo y no como sustituto del juicio experto, requiriéndose validación humana, trazabilidad y auditoría periódica.

LIMITACIONES DEL ESTUDIO

La base real disponible para modelado es acotada y potencialmente no representa toda la variabilidad de expedientes del sector minero.

La calidad y completitud de los datos históricos condiciona el aprendizaje del modelo y puede trasladar sesgos previos al sistema automatizado.

El uso de datos sintéticos, aunque útil para ampliar cobertura y balancear clases, no sustituye la necesidad de validación externa con datos reales independientes.

Finalmente, los hallazgos no son transferibles de forma automática a otros sectores ambientales, por lo que se requiere evaluación específica antes de su adopción en nuevos contextos.

CONCLUSIONES

El modelo Random Forest, con su mayor capacidad de interpretabilidad y estabilidad, parece ser el más adecuado para el sector ambiental, ya que facilita la toma de decisiones informada y la reducción de errores humanos en un área crítica como la evaluación de estudios ambientales.

Los resultados obtenidos hasta el momento mejoran la capacidad predictiva de los sistemas de evaluación ambiental automatizada usando un modelo híbrido que integra reglas expertas ambientales con algoritmos de aprendizaje automático como lo es el modelo Random Forest, complementado con datos sintéticos ambientalmente coherentes y una arquitectura que permite predecir resultados.

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a mi asesor, Miguel Herrero Uceda, por su guía y asesoría invaluable, su retroalimentación es fundamental para el desarrollo de este proyecto y para mi crecimiento académico y profesional.

Así también, agradecer a la Revista de Investigación Multidisciplinaria Iberoamericana (RIMI) por darme la oportunidad de compartir este proyecto de tesis con la comunidad académica.

REFERENCIAS

- European Environment Agency. (2023). AI4EIA Project: Interim results. *EEA Technical Report, No 15/2023*.
- Liu, H. (2023). Transformer-based analysis of environmental documents. *AI for Earth Sciences*, 7(1), 45-60.
- Kumar, P. (2024). *Large language models (LLMs): Survey, technical frameworks, and future challenges*. Artificial Intelligence Review. <https://doi.org/10.1007/s10462-024-10888-y>
- Monteagudo, L. A. (2023, 08 30). *Ideas medioambientales*. Retrieved from <https://ideasmedioambientales.com/>: <https://ideasmedioambientales.com/inteligencia-artificial-evaluacion-de-impacto-ambiental/>
- Mosqueira-Rey, J., et al. (2023). *Human-in-the-loop machine learning: A state of the art*. Artificial Intelligence Review. <https://doi.org/10.1007/s10462-022-10246-w>
- OECD. (2023). *Artificial Intelligence in Government: IA4Gov Report*. OECD Digital Government Studies. Obtenido de <https://www.oecd.org/gov/digital-government/artificial-intelligence-in-government.htm>

- Tremblay, J. (2023). Canada's AI-powered environmental assessment system. *Environmental Policy and Governance*, 33(4), 345-360.
- Zambrano Salazar, L. A. (2024). Aplicación de la Inteligencia Artificial en la evaluación del impacto ambiental de proyectos de ingeniería civil. *MAGAZINE DE LAS CIENCIAS*, 114. doi:<https://doi.org/10.33262/rmc.v9i3.3163>
- Zubiaga, A. (2024). *Natural language processing in the era of large language models*. Frontiers in Artificial Intelligence. <https://doi.org/10.3389/frai.2023.1350306>