Artículo de investigación https://doi.org/10.47460/minerva.v7i21.372

Evaluación del impacto de la inteligencia artificial en la calidad de la información contable y la eficiencia de las decisiones organizacionales

Elaine Sandra Monterroso Achahui*
https://orcid.org/0009-0004-3360-2224
elaine.monterroso@gmail.com
Universidad Nacional San Antonio Abad del Cusco
Cusco, Perú

Magaly Pezo Zegarra
https://orcid.org/0009-0002-8549-5251
mpezozegarra@gmail.com
Universidad Nacional San Antonio Abad del Cusco
Cusco, Perú

Autora de correspondencia: elaine.monterroso@gmail.com

Recibido: (11/08/2026), Aceptado: (31/10/2026)

Resumen. La investigación tuvo como objetivo evaluar el desempeño de estrategias de inteligencia artificial en la detección de anomalías contables y su contribución a la eficiencia de los procesos de revisión. Se desarrolló un estudio cuantitativo experimental con registros contables sintéticos correspondientes a dos ejercicios fiscales. Se compararon cuatro estrategias: reglas tradicionales, Isolation Forest, Random Forest y un modelo híbrido, mediante indicadores de precisión promedio, área bajo la curva ROC y cobertura de anomalías bajo diferentes presupuestos de inspección. Random Forest obtuvo el mejor desempeño, con una precisión promedio de 0,5392 y un área ROC de 0,9517. Con un presupuesto de revisión del 2 %, identificó 379 de 427 anomalías (88,76 %). Los resultados evidencian el potencial de la inteligencia artificial para priorizar registros contables, sin embargo, se sugiere su aplicación organizacional luego de haberse verificado empíricamente.

Palabras clave: inteligencia artificial; información contable; detección de anomalías; aprendizaje automático.

Assessing the Impact of Artificial Intelligence on the Quality of Accounting Information and the Efficiency of Organizational Decisions

Abstract. This study aimed to evaluate the performance of artificial intelligence strategies in detecting accounting anomalies and their contribution to the efficiency of review processes. A quantitative experimental study was conducted using synthetic accounting records from two fiscal years. Four strategies were compared: traditional rules, Isolation Forest, Random Forest, and a hybrid model. Performance was assessed using average precision, area under the ROC curve, and anomaly coverage under different inspection budgets. Random Forest achieved the best overall performance, with an average precision of 0.5392 and a ROC-AUC of 0.9517. Under a 2% review budget, it identified 379 out of 427 anomalies (88.76%). The results show the potential of artificial intelligence to prioritize accounting records, however, its organizational application is suggested after having been empirically verified.

Keywords: artificial intelligence; accounting information; anomaly detection; machine learning.

I. INTRODUCCIÓN

La transformación digital de los procesos contables ha incrementado la disponibilidad de información financiera y las posibilidades de automatizar su procesamiento, análisis y verificación. Sin embargo, el crecimiento del volumen de transacciones también plantea dificultades para identificar oportunamente registros inconsistentes, operaciones atípicas y patrones que requieren revisión. Desde la perspectiva de la información financiera, la utilidad de los datos contables depende de características como la relevancia, la representación fiel, la comparabilidad, la verificabilidad y la oportunidad [1]. En este contexto, los sistemas de información contable desempeñan un papel fundamental en la generación de información que respalde los procesos de control y las decisiones organizacionales [2].

La incorporación de inteligencia artificial en los procedimientos contables ha ampliado las posibilidades de analizar grandes conjuntos de transacciones mediante algoritmos capaces de reconocer patrones y detectar comportamientos inusuales. Kokina y Davenport [3] identificaron la automatización inteligente como una transformación relevante para las actividades de auditoría, mientras que Appelbaum et al. [4] destacaron la importancia de integrar procedimientos analíticos en los procesos de revisión financiera.

Las investigaciones sobre detección computacional de irregularidades financieras han mostrado que los algoritmos de aprendizaje automático pueden aportar capacidades analíticas adicionales a los procedimientos estadísticos convencionales. Perols [5] comparó métodos estadísticos y de aprendizaje automático para identificar fraude en estados financieros, mientras que Bao et al. [6] desarrollaron un enfoque predictivo para detectar fraude contable en empresas cotizadas. No obstante, la identificación de anomalías en asientos contables representa un problema diferente, debido a la estructura transaccional de los registros, la heterogeneidad de las operaciones y la baja frecuencia relativa de los casos anómalos.

En los procesos de control contable, la disponibilidad de recursos de inspección constituye una restricción operativa relevante. Un sistema capaz de identificar anomalías puede resultar poco útil si genera una cantidad excesiva de alertas o si las operaciones de mayor interés quedan fuera de las prioridades de revisión. Por esta razón, la evaluación de los algoritmos debe incorporar indicadores que relacionen el desempeño predictivo con la proporción de registros examinados. La literatura metodológica ha señalado que las curvas precisión-recall proporcionan información especialmente relevante cuando las clases presentan distribuciones desequilibradas [7], [8]. Este enfoque permite analizar la utilidad de los modelos bajo diferentes capacidades de inspección y establecer comparaciones más próximas a las necesidades de priorización contable.

En este contexto, el objetivo de la investigación fue evaluar el desempeño de diferentes estrategias de inteligencia artificial en la identificación de anomalías contables y determinar su contribución a la eficiencia de las decisiones de revisión mediante un diseño experimental computacional. Se compararon procedimientos basados en reglas tradicionales, Isolation Forest, Random Forest y una estrategia híbrida, utilizando registros correspondientes a dos ejercicios fiscales consecutivos. La contribución del estudio consiste en integrar la evaluación predictiva con indicadores de cobertura y capacidad de inspección, así como examinar las diferencias de desempeño entre categorías de anomalías. De esta manera, se busca establecer evidencia cuantitativa sobre las condiciones en las que la inteligencia artificial puede mejorar la priorización de registros contables y las limitaciones que deben considerarse para su aplicación en sistemas de apoyo a las decisiones organizacionales.

II. MARCO TEÓRICO

A. Calidad de la información contable y utilidad para las decisiones organizacionales

La calidad de la información contable constituye un componente fundamental de los sistemas de gestión, debido a su función en la representación de las operaciones económicas y en el suministro de información para los procesos de control, evaluación y toma de decisiones. El marco conceptual de información financiera establece que la relevancia y la representación fiel son características fundamentales de la información útil, mientras que la comparabilidad, verificabilidad, oportunidad y comprensibilidad contribuyen a incrementar su utilidad [1]. Estas características permiten comprender la calidad contable como un concepto multidimensional que no depende exclusivamente de la exactitud de los registros, sino también de su capacidad para representar adecuadamente los fenómenos económicos.

Los sistemas de información contable integran procedimientos de registro, clasificación, procesamiento y comunicación de datos financieros. Su contribución a la gestión organizacional depende de la confiabilidad de los registros y de la posibilidad de transformar los datos disponibles en información pertinente para las decisiones. La calidad de estos sistemas se relaciona con la utilidad de la información producida y con las condiciones en que los responsables organizacionales pueden utilizarla para evaluar alternativas [2].

La presencia de operaciones atípicas, inconsistencias o registros que requieren verificación puede afectar la confiabilidad de los procesos contables y aumentar las necesidades de revisión. Sin embargo, no todas las anomalías representan errores materiales, incumplimientos normativos o comportamientos fraudulentos. Una operación puede presentar características estadísticas poco frecuentes y, al mismo tiempo, corresponder a una transacción legítima. En consecuencia, los procedimientos de detección deben interpretarse como mecanismos de identificación de riesgos y selección de registros para análisis posterior, sin sustituir la valoración profesional de su naturaleza y consecuencias.

B. Inteligencia artificial y detección de anomalías en registros contables

La inteligencia artificial ha ampliado las posibilidades de automatización de los procesos de auditoría y análisis contable mediante herramientas capaces de procesar grandes volúmenes de información e identificar relaciones difíciles de establecer a través de procedimientos manuales. Kokina y Davenport [3] examinaron las implicaciones de las tecnologías cognitivas en las actividades de auditoría, destacando su potencial para modificar la distribución de tareas entre sistemas automatizados y profesionales. De manera complementaria, Appelbaum et al. [4] sistematizaron las aplicaciones de procedimientos analíticos en auditoría externa y propusieron un marco para comprender su integración en actividades de evaluación financiera.

La detección de anomalías constituye una aplicación particularmente relevante del aprendizaje automático en el ámbito contable. Su propósito consiste en identificar observaciones que presentan diferencias respecto de patrones considerados habituales. Estas diferencias pueden relacionarse con importes, frecuencias, fechas, cuentas involucradas, usuarios responsables o combinaciones de características operativas. La interpretación de una anomalía requiere considerar el contexto contable, debido a que la singularidad estadística no determina necesariamente la existencia de una irregularidad financiera.

Los métodos tradicionales de revisión suelen fundamentarse en reglas previamente establecidas, como límites monetarios, operaciones registradas fuera de horarios habituales o diferencias temporales entre la documentación y la contabilización. Estas estrategias ofrecen criterios comprensibles y pueden resultar útiles cuando los riesgos se encuentran claramente definidos. Sin embargo, su capacidad para identificar relaciones complejas entre múltiples variables puede ser limitada. Perols [5] examinó diferentes algoritmos estadísticos y de aprendizaje automático para detectar fraude en estados financieros, evidenciando que el desempeño depende de las características de los modelos y de las condiciones de evaluación.

Los algoritmos supervisados utilizan observaciones previamente clasificadas para aprender relaciones entre las características de entrada y las categorías de interés. Bao et al. [6] aplicaron técnicas de aprendizaje automático a la identificación de fraude contable en empresas cotizadas, proporcionando evidencia sobre las posibilidades de estos procedimientos para complementar el análisis financiero. Entre los algoritmos supervisados, Random Forest combina múltiples árboles de decisión y agrega sus predicciones para obtener resultados de clasificación [9]. Su estructura permite representar relaciones no lineales e interacciones entre variables sin requerir una especificación funcional única.

Por su parte, los métodos no supervisados identifican patrones atípicos sin depender de etiquetas de anomalía durante el entrenamiento. Isolation Forest utiliza particiones aleatorias para aislar observaciones, bajo el principio de que los casos anómalos pueden requerir menos divisiones para separarse del conjunto de datos [10]. Esta característica lo convierte en una alternativa para escenarios en los que las irregularidades conocidas son escasas o no existe una clasificación previa suficientemente confiable.

Los enfoques híbridos combinan diferentes procedimientos con el propósito de aprovechar sus características complementarias. No obstante, su rendimiento depende de la compatibilidad entre las puntuaciones, los criterios de integración y las propiedades de los datos. La combinación de modelos no garantiza una mejora automática, especialmente cuando sus errores presentan patrones semejantes o cuando alguno de los componentes incorpora información poco discriminativa.

En el análisis de registros contables, la estructura de los asientos introduce consideraciones adicionales. Cada operación puede comprender varias líneas relacionadas mediante cuentas de débito y crédito, por lo que su evaluación requiere preservar la unidad transaccional. Gronewald et al. [11] desarrollaron un conjunto de datos sintéticos que reproduce esta estructura e incorpora categorías identificadas de anomalías, permitiendo evaluar diferentes estrategias de detección bajo condiciones experimentales controladas.

C. Desempeño predictivo y eficiencia de la revisión contable

La evaluación de algoritmos de detección de anomalías requiere considerar tanto su capacidad para distinguir registros normales y anómalos como su utilidad dentro de procesos de revisión con recursos limitados. En conjuntos de datos con baja proporción de anomalías, la exactitud global puede ofrecer una interpretación insuficiente del desempeño, debido a que un clasificador puede alcanzar valores elevados al identificar correctamente la clase mayoritaria sin reconocer adecuadamente los casos de interés.

Las medidas de precisión y sensibilidad permiten examinar aspectos complementarios del rendimiento. La precisión expresa la proporción de anomalías verdaderas entre los registros seleccionados como positivos, mientras que la sensibilidad representa la proporción de anomalías existentes que fueron identificadas. Davis y Goadrich [7] estudiaron la relación entre las curvas precisión-recall y ROC, estableciendo fundamentos para interpretar estas representaciones del desempeño. Posteriormente, Saito y Rehmsmeier [8] destacaron la utilidad de las curvas precisión-recall en conjuntos de datos con distribuciones desequilibradas.

En los procesos de revisión contable, la capacidad de inspección puede expresarse como una proporción del total de operaciones disponibles. Bajo esta condición, los modelos no solamente deben identificar anomalías, sino también ordenarlas según su prioridad estimada. La selección de los registros con mayores puntuaciones permite analizar cuántas anomalías pueden identificarse cuando la organización dispone de recursos para examinar únicamente una fracción de las operaciones.

La evaluación por categorías de anomalía proporciona información adicional sobre las limitaciones de los modelos. Un algoritmo puede presentar un desempeño global elevado y, simultáneamente, dificultades para identificar determinados tipos de registros. Por ello, el análisis desagregado resulta necesario para establecer la consistencia de los procedimientos y evitar que los indicadores agregados oculten deficiencias relevantes para el control contable.

III. METODOLOGÍA

A. Diseño de la investigación

La investigación se desarrolló mediante un enfoque cuantitativo, de carácter experimental computacional, orientado a evaluar el desempeño de diferentes métodos de identificación de anomalías en registros contables y su contribución a la eficiencia de los procesos de revisión. Se estableció un diseño comparativo que incluyó procedimientos basados en reglas tradicionales, algoritmos de aprendizaje automático y una estrategia híbrida de priorización de registros.

La evaluación se fundamentó en la capacidad de los métodos para identificar operaciones contables previamente clasificadas como anómalas, considerando tanto el rendimiento predictivo como la cantidad de registros que debían examinarse para alcanzar determinados niveles de detección. La eficiencia de las decisiones se operacionalizó mediante la selección priorizada de asientos para revisión, bajo restricciones de capacidad de inspección.

B. Fuente de datos y unidad de análisis

Se utilizó el conjunto de datos DIB General Ledger Data, publicado en Zenodo, compuesto por registros contables sintéticos correspondientes a dos ejercicios fiscales consecutivos, 2023 y 2024. El conjunto reproduce operaciones de partida doble e incorpora diferentes categorías de anomalías identificadas mediante etiquetas que permiten evaluar los procedimientos de detección.

La base original contenía 256 232 líneas contables, agrupadas en 102 429 asientos únicos. De estos, 101 579 correspondían a operaciones clasificadas como normales y 850 presentaban anomalías. La proporción global de asientos anómalos fue del 0,83 %. La unidad de análisis fue el asiento contable, definido como el conjunto de líneas asociadas a un mismo identificador de registro. Esta decisión

permitió conservar la estructura transaccional de las operaciones y evitar que las líneas pertenecientes a un mismo asiento fueran tratadas como observaciones independientes.

La información se organizó en una matriz analítica que incorporó características temporales, monetarias y operativas de cada asiento. Se incluyeron el número de líneas, la cantidad de cuentas involucradas, el importe de la operación, la tasa impositiva máxima registrada, el mes y día de contabilización, la hora de registro, el tiempo transcurrido entre la fecha documental y el ingreso de la operación, y la identificación del usuario responsable.

C. Preparación y verificación de los datos

El procesamiento inicial comprendió la lectura de los registros originales, su agrupación por identificador de asiento y la construcción de variables agregadas. Se verificó la integridad de los campos, la consistencia de las etiquetas y el equilibrio entre los importes registrados al debe y al haber. No se identificaron valores ausentes en las 19 variables originales. Asimismo, la comprobación de los asientos agrupados no identificó diferencias entre débitos y créditos superiores a una tolerancia de 0,01 unidades monetarias.

Para reducir la asimetría de los importes, se aplicó una transformación logarítmica a la magnitud monetaria de las operaciones. Las variables categóricas correspondientes a los usuarios fueron codificadas mediante representación binaria, mientras que los valores numéricos se incorporaron al proceso de modelado conservando sus respectivas escalas. Las variables que identificaban directamente las anomalías, incluidas las etiquetas de clasificación y sus categorías, se excluyeron de los atributos predictivos. También se excluyeron indicadores derivados del mecanismo de generación de anomalías, con el propósito de reducir el riesgo de incorporar información sobre la respuesta esperada durante el entrenamiento.

D. Diseño experimental y modelos de comparación

Se estableció una partición temporal de los datos. Los 51 064 asientos correspondientes a 2023, entre ellos 423 anomalías, se utilizaron para el entrenamiento de los modelos y la determinación de los parámetros de referencia. Los 51 365 asientos de 2024, que incluían 427 anomalías, se reservaron para la evaluación. Se implementaron cuatro estrategias de identificación y priorización. El primer método consistió en un sistema de reglas tradicionales que asignó puntuaciones de riesgo según retrasos de registro, importes elevados, operaciones realizadas fuera del horario habitual, contabilizaciones durante fines de semana y tasas impositivas diferentes del valor de referencia. Los umbrales correspondientes al retraso y al importe se establecieron utilizando el percentil 95 de los datos de entrenamiento.

El segundo método utilizó Isolation Forest, un algoritmo no supervisado de detección de anomalías, configurado con 150 estimadores y una semilla aleatoria de 42. Las puntuaciones de anomalía obtenidas se utilizaron para ordenar los asientos según su prioridad de revisión. El tercer método empleó Random Forest, mediante un modelo supervisado de clasificación con 180 árboles, un mínimo de tres observaciones por hoja y ponderación de clases para considerar el desequilibrio entre registros normales y anómalos. El modelo generó una puntuación predictiva para cada asiento del período de prueba.

El cuarto método integró las puntuaciones de Random Forest y del sistema de reglas mediante una combinación ponderada de sus posiciones relativas en la clasificación. Se utilizaron pesos de 0,80 y 0,20, respectivamente, para explorar el comportamiento de una estrategia híbrida de priorización.

E. Indicadores de desempeño y eficiencia

El desempeño de los modelos se evaluó mediante el área bajo la curva precisión-recall, el área bajo la curva ROC y los indicadores de precisión y sensibilidad calculados para diferentes capacidades de revisión. Se establecieron presupuestos de inspección equivalentes al 1 %, 2 %, 5 % y 10 % de los asientos del conjunto de prueba. Para cada presupuesto se seleccionaron los registros con mayores puntuaciones de riesgo y se contabilizó la cantidad de anomalías correctamente identificadas. La precisión de revisión se definió como:

$$P_k = \frac{TP_k}{k} \tag{1}$$

donde TP_k representa el número de asientos anómalos identificados entre los k registros seleccionados para revisión. La cobertura de anomalías se calculó mediante:

$$R_k = \frac{TP_k}{N_a} \tag{2}$$

donde N_a corresponde al número total de asientos anómalos presentes en el conjunto de prueba.

Adicionalmente, se examinó el comportamiento de los métodos en las ocho categorías de anomalías disponibles, con el propósito de identificar diferencias en su capacidad de detección y posibles limitaciones de las estrategias de priorización.

F. Alcance de la evaluación

El diseño permitió evaluar la capacidad de diferentes procedimientos computacionales para identificar y priorizar registros contables anómalos, utilizando un conjunto de datos con etiquetas conocidas. La calidad de la información contable se examinó desde la perspectiva de la detección de irregularidades y de la utilidad de las alertas generadas para orientar procesos de revisión. La eficiencia de las decisiones organizacionales se representó mediante la asignación de capacidades limitadas de inspección a los registros considerados de mayor riesgo.

Dado el carácter sintético del conjunto de datos, los resultados corresponden a un entorno experimental y no constituyen mediciones directas de productividad, calidad integral de los estados financieros ni decisiones efectivamente adoptadas por organizaciones reales.

IV. RESULTADOS

A. Desempeño comparativo de los modelos

El conjunto de prueba comprendió 51 365 asientos contables, de los cuales 427 (0,83 %) estaban etiquetados como anómalos. La evaluación de los cuatro procedimientos mostró diferencias en la capacidad para ordenar los registros según su riesgo de anomalía. Random Forest alcanzó la mayor precisión promedio (AP = 0,5392) y el mayor AUC-ROC (0,9517), seguido por el modelo híbrido (AP = 0,4300; AUC-ROC = 0,9449). Los procedimientos basados en reglas tradicionales y en Isolation Forest obtuvieron valores AP de 0,0434 y 0,0540, respectivamente. La tabla 1 presenta los indicadores predictivos y la cantidad de anomalías identificadas cuando se inspecciona el 2 % de los asientos. La Fig. 1 permite comparar visualmente la precisión promedio de los cuatro métodos.

Tabla 1. Desempeño comparativo de los modelos de detección de anomalías contables.

Modelo	AP	AUC-ROC	Detectadas al 2 %
Reglas tradicionales	0.0434	0.7786	51
Isolation Forest	0.0540	0.8656	87
Random Forest	0.5392	0.9517	379
Híbrido (RF + reglas)	0.4300	0.9449	270

La Fig. 1 permite visualizar las diferencias en la precisión promedio alcanzada por los modelos evaluados, destacando el desempeño del algoritmo Random Forest frente a los métodos tradicionales y no supervisados.

La estrategia híbrida no superó al clasificador supervisado independiente. Este comportamiento indica que la incorporación de una puntuación adicional basada en reglas no mejoró la clasificación obtenida con la configuración ensayada.

B. Eficiencia de la priorización de revisiones

La eficiencia de la selección se examinó mediante presupuestos de revisión equivalentes al 1 %, 2 %, 5 % y 10 % de los asientos de prueba. En el presupuesto del 2 %, Random Forest identificó 379 de las 427 anomalías (88,76 %), frente a 51 (11,94 %) para las reglas tradicionales, 87 (20,37 %) para Isolation Forest y 270 (63,23 %) para el modelo híbrido. La diferencia de cobertura entre Random Forest y las reglas fue de 76,81 puntos porcentuales. La tabla 2 resume el número de anomalías recuperadas según la capacidad de revisión. El aumento del presupuesto no produjo incrementos uniformes: Random

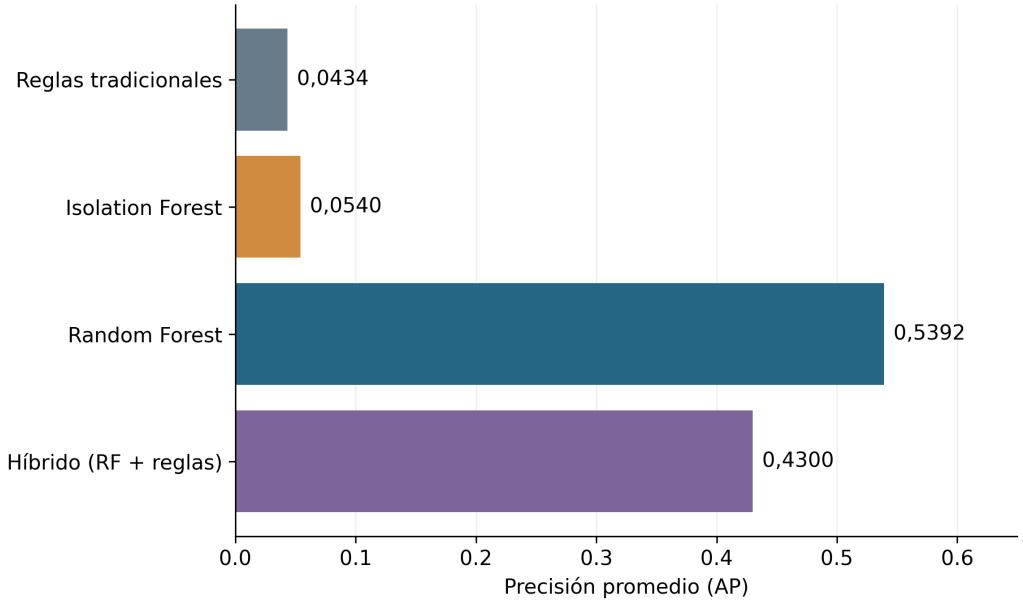


Fig. 1. Precisión promedio según el tipo de modelo empleado

Forest recuperó 386 anomalías con el 5 % y 390 con el 10 %, lo que sugiere rendimientos marginales decrecientes dentro de este conjunto de prueba.

Tabla 2. Anomalías detectadas según el presupuesto de revisión contable.

Modelo	**1 %**	**2 %**	**5 %**	**10 %**
Reglas tradicionales	36	51	105	275
Isolation Forest	46	87	163	241
Random Forest	277	379	386	390
Híbrido (RF + reglas)	250	270	386	390

La Fig. 2 muestra la evolución de la cobertura de anomalías en función del presupuesto de inspección, permitiendo comparar la eficiencia de las estrategias de priorización y determinar cuáles concentran una mayor proporción de registros anómalos con un menor esfuerzo de revisión.

Los indicadores representan eficiencia en la priorización experimental de registros y no una reducción observada de tiempos o costos de auditoría en una organización real.

C. Comportamiento por categoría de anomalía

La capacidad de identificación no fue homogénea entre las ocho categorías de anomalías. Al revisar el 2 % de los asientos, Random Forest recuperó la totalidad de los casos de las categorías 1, 2, 4, 5 y 6, e identificó 54 de los 60 casos de la categoría 3. En cambio, no recuperó casos de las categorías 7 y 8, que reunían 30 y 12 anomalías, respectivamente.

La tabla 3 detalla los resultados por categoría y la Fig. 3 compara las tasas de detección de los modelos. Esta heterogeneidad limita la interpretación de un indicador agregado elevado como evidencia de desempeño uniforme ante todas las clases de irregularidad.

Para profundizar en las diferencias observadas entre los modelos, se analizó su capacidad de detección según las categorías de anomalías contables identificadas en el conjunto de datos. La Fig. 3 presenta las tasas de detección correspondientes a cada categoría, permitiendo reconocer los tipos de irregularidades que fueron identificados con mayor eficacia y aquellos que representaron mayores dificultades para los algoritmos. Este análisis proporciona una perspectiva más detallada del desempeño de los modelos y permite identificar limitaciones relevantes para su aplicación en los procesos de revisión contable.

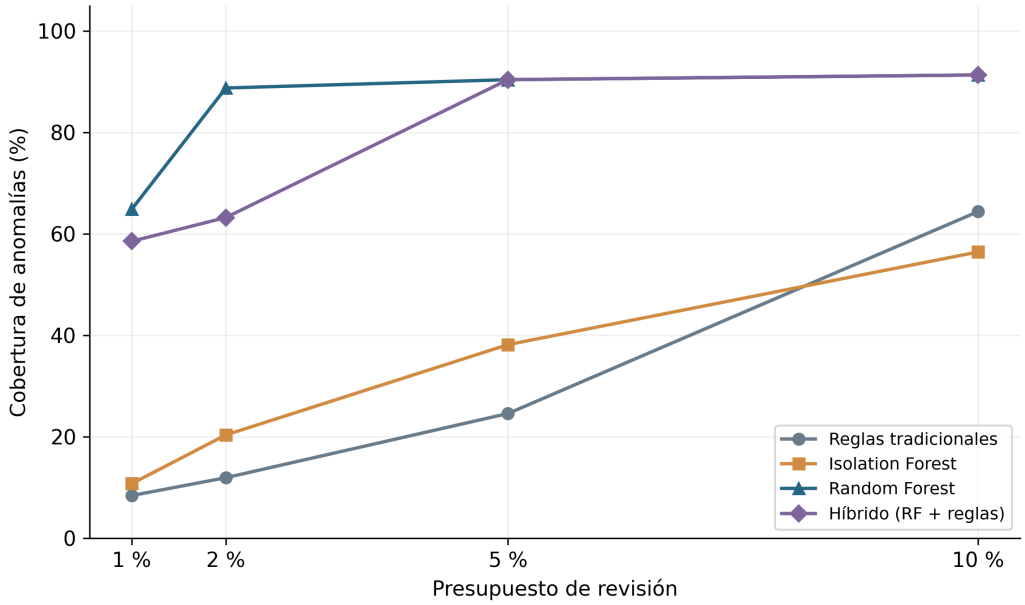


Fig. 2. Cobertura según el presupuesto

Tabla 3. Desempeño de los modelos según la categoría de anomalía contable.

Categoría	Casos	Reglas tradicionales	Isolation Forest	Random Forest	Híbrido (RF + reglas)
1	60	20	10	60	60
2	30	3	7	30	29
3	60	2	15	54	13
4	60	4	34	60	60
5	50	8	9	50	47
6	125	13	12	125	61
7	30	0	0	0	0
8	12	1	0	0	0

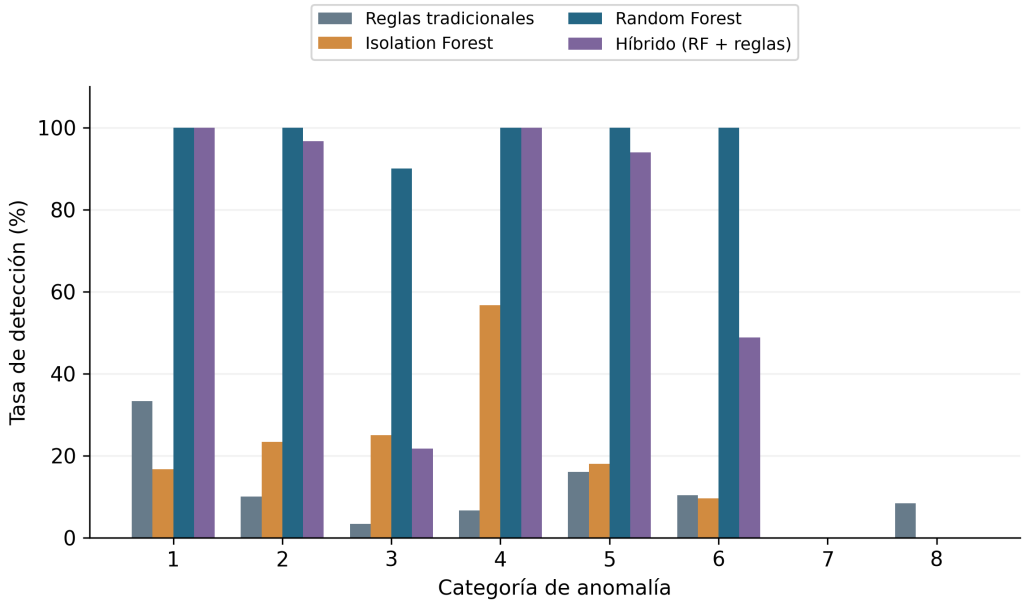


Fig. 3. Detección por categoría

D. Incertidumbre de la precisión promedio

Se estimaron intervalos percentiles del 95 % para la precisión promedio mediante 400 remuestreos bootstrap del conjunto de prueba. Random Forest presentó un intervalo de 0,4907 a 0,5907; el híbrido, de 0,3834 a 0,4798; Isolation Forest, de 0,0468 a 0,0642; y las reglas tradicionales, de 0,0371 a 0,0520. La tabla 4 presenta los valores puntuales y sus intervalos.

Tabla 4. Precisión promedio e intervalos de confianza del 95 % de los modelos evaluados.

Modelo	AP	IC 95 % inferior	IC 95 % superior
Reglas tradicionales	0.0434	0.0371	0.0520
Isolation Forest	0.0540	0.0468	0.0642
Random Forest	0.5392	0.4907	0.5907
Híbrido (RF + reglas)	0.4300	0.3834	0.4798

Estos intervalos describen la variabilidad del indicador ante el remuestreo de las observaciones de prueba, condicionada a los modelos ya entrenados. No cuantifican la incertidumbre del reentrenamiento ni permiten extrapolar directamente los resultados a registros contables reales.

E. Discusión

Los resultados evidenciaron que Random Forest presentó un desempeño superior al de los procedimientos tradicionales, Isolation Forest y el enfoque híbrido, alcanzando una precisión promedio de 0,5392 y una cobertura del 88,76 % de las anomalías con un presupuesto de revisión del 2 %. Este comportamiento es consistente con el potencial de los algoritmos supervisados para identificar patrones complejos en información financiera, señalado por Perols [5] y Bao et al. [6]. Asimismo, la diferencia entre los indicadores ROC-AUC y precisión promedio confirma la importancia de emplear métricas complementarias cuando las anomalías representan una proporción reducida del conjunto de datos, como plantean Davis y Goadrich [7] y Saito y Rehmsmeier [8].

La evaluación por categorías mostró diferencias en la capacidad de identificación de los modelos, lo que indica que un desempeño global favorable no garantiza resultados uniformes frente a todas las irregularidades. Además, la estrategia híbrida no superó a Random Forest, evidenciando que la combinación de procedimientos no asegura mejoras predictivas. Estos hallazgos resaltan la necesidad de evaluar los algoritmos considerando tanto su capacidad discriminativa como la cobertura alcanzada bajo restricciones de inspección, especialmente cuando su finalidad es orientar la selección de registros contables para revisión.

Los resultados deben interpretarse dentro de las condiciones del conjunto de datos sintéticos utilizado [11]. Aunque el experimento permitió comparar estrategias bajo condiciones reproducibles, no demuestra directamente mejoras en la calidad integral de la información financiera ni en los resultados de decisiones organizacionales reales. Su principal contribución consiste en aportar evidencia cuantitativa sobre la eficiencia de la priorización de anomalías contables mediante inteligencia artificial, cuya transferencia a entornos operativos requiere validaciones adicionales.

CONCLUSIONES

La investigación permitió evaluar comparativamente el desempeño de cuatro estrategias de detección de anomalías contables y su contribución a la priorización de registros para revisión. Los resultados evidenciaron diferencias significativas en términos operativos entre los procedimientos basados en reglas, Isolation Forest, Random Forest y el enfoque híbrido, destacando la importancia de seleccionar los algoritmos según su capacidad de identificación y las restricciones de inspección disponibles.

Random Forest presentó el mejor desempeño global, con un área bajo la curva ROC de 0,9517 y una precisión promedio de 0,5392. Bajo un presupuesto de revisión equivalente al 2 % de los asientos contables, permitió identificar 379 de las 427 anomalías presentes en el conjunto de evaluación, correspondientes a una cobertura del 88,76 %. Estos resultados muestran que la clasificación supervisada

puede concentrar una proporción elevada de operaciones anómalas dentro de un grupo reducido de registros, favoreciendo la eficiencia de los procedimientos de revisión.

El análisis por categorías evidenció que un rendimiento global elevado no garantiza una identificación uniforme de todas las irregularidades. Las diferencias observadas entre tipos de anomalías confirman la necesidad de complementar los indicadores agregados con evaluaciones desagregadas, particularmente cuando determinados patrones presentan mayores dificultades de reconocimiento.

La comparación de los modelos también permitió establecer que la incorporación de estrategias híbridas no produce necesariamente mejoras respecto de los algoritmos individuales. Su utilidad depende de los criterios de integración y de la capacidad complementaria de los procedimientos combinados, por lo que su aplicación requiere validación experimental específica.

Los hallazgos respaldan el potencial de la inteligencia artificial como herramienta de apoyo a los procesos de control contable, especialmente mediante la priorización de operaciones que requieren verificación. No obstante, al utilizar registros sintéticos, las conclusiones se circunscriben al entorno experimental evaluado. La comprobación de beneficios sobre la calidad integral de la información financiera y la eficiencia de las decisiones organizacionales requiere estudios posteriores con datos reales, participación de profesionales contables e indicadores operativos y económicos de desempeño.

REFERENCIAS

[1] International Accounting Standards Board, *Conceptual Framework for Financial Reporting*. London, U.K.: IFRS Foundation, 2018. [Online]. Available: https://www.ifrs.org/issued-standards/list-of-standards/conceptual-framework/

[2] N. T. T. Thuy, "Effect of accounting information system quality on decision-making success and non-financial performance: Does non-financial information quality matter?" *Cogent Business & Management*, vol. 12, no. 1, 2025, art. no. 2447913, doi: 10.1080/23311975.2024.2447913.

[3] J. Kokina and T. H. Davenport, "The emergence of artificial intelligence: How automation is changing auditing," *Journal of Emerging Technologies in Accounting*, vol. 14, no. 1, pp. 115–122, 2017, doi: 10.2308/jeta-51730.

[4] D. A. Appelbaum, A. Kogan, and M. A. Vasarhelyi, "Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics," *Journal of Accounting Literature*, vol. 40, pp. 83–101, 2018, doi: 10.1016/j.acclit.2018.01.001.

[5] J. Perols, "Financial statement fraud detection: An analysis of statistical and machine learning algorithms," *Auditing: A Journal of Practice & Theory*, vol. 30, no. 2, pp. 19–50, 2011, doi: 10.2308/ajpt-50009.

[6] Y. Bao, B. Ke, B. Li, Y. J. Yu, and J. Zhang, "Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach," *Journal of Accounting Research*, vol. 58, no. 1, pp. 199–235, 2020, doi: 10.1111/1475-679X.12292.

[7] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, USA, 2006, pp. 233–240, doi: 10.1145/1143844.1143874.

[8] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, 2015, art. no. e0118432, doi: 10.1371/journal.pone.0118432.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[10] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of the 8th IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.

[11] J. Gronewald, S. Stephan, A. Rombach, and P. Fettke, "Generating synthetic journal entries for audit analytics: Method and dataset," *Data in Brief*, vol. 69, 2026, art. no. 113257, doi: 10.1016/j.dib.2026.113257.

[12] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, 2012, art. no. 3, doi: 10.1145/2133360.2133363.

[13] M. Schreyer, T. Sattarov, D. Borth, A. Dengel, and B. Reimer, "Detection of anomalies in large scale accounting data using deep autoencoder networks," arXiv:1709.05254, 2017, doi: 10.48550/arXiv.1709.05254.

[14] J. Gronewald, S. Stephan, A. Rombach, and P. Fettke, "DIB General Ledger Data," Zenodo, 2026, doi: 10.5281/zenodo.22014653.