

# Aplicación del Modelo de Machine Learning (Random Forest) para la Identificación de los Indicadores Financieros más Influyentes en la Predicción del Desempeño de la Industria Minera Peruana

## Introducción

Uno de los sectores económicos más relevantes de la economía peruana, es el minero, que aporta significativamente al Producto Bruto Interno (PBI) del país, representado con más del 57% del valor de las exportaciones y con una fuerte contribución dentro los ingresos fiscales (Lizarzaburu, Noriega, Macías, Rodríguez, y Berggrun, 2019). El periodo 2020-2024 destaca por su alta volatilidad, marcada por la contracción generada por la pandemia de COVID-19 en 2020 (Velasquez, Zegarra, y Fernandez, 2021) y la posterior recuperación impulsada por el “superciclo” de precios de los commodities desde 2021 (BBVA Research, 2023).

Este escenario de crisis y auge constituye un entorno propicio para analizar los factores financieros que determinan la resiliencia empresarial. Sin embargo, los modelos estadísticos tradicionales presentan limitaciones para capturar las relaciones no lineales del mercado. Frente a ello, los algoritmos de Machine Learning como Random Forest ofrecen una alternativa para identificar patrones complejos sin supuestos restrictivos (Sun y Deng, 2025). En este contexto, el objetivo del presente estudio es aplicar el algoritmo Random Forest para determinar la importancia relativa de los ratios de liquidez, rentabilidad, gestión y endeudamiento en la predicción del desempeño de las empresas mineras peruanas durante el periodo 2020–2024.

## El algoritmo Random Forest: un enfoque de ensamblaje

El algoritmo Random Forest fue introducido por Leo Breiman (2001), quien menciona que el mismo pertenece a los métodos de ensamblaje (*ensemble learning*) y se basa en combinar múltiples árboles de decisión para obtener predicciones más precisas y estables. Cada árbol se entrena con una muestra aleatoria de los datos mediante el método *bagging*, y en cada división se selecciona aleatoriamente un subconjunto de variables predictoras, lo que reduce la correlación entre árboles y evita el sobreajuste (Hastie, Tibshirani, y Friedman, 2009). En tareas de clasificación, como la de este estudio, la predicción final se determina por votación mayoritaria entre todos los árboles del bosque. Esta estructura permite al modelo capturar relaciones no lineales y evaluar la importancia relativa de cada variable, convirtiéndolo en una herramienta eficaz para el análisis financiero y la predicción del desempeño empresarial (Goodfellow, Bengio, y Courville, 2018).

## Medición de la importancia de variables (*Feature Importance*)

Una de las principales ventajas del algoritmo Random Forest es su capacidad para estimar la relevancia de cada variable predictor en el modelo (James, Witten, Hastie, y Tibshirani, 2013). Entre los métodos disponibles, destacan la Importancia Gini (MDI) y la Importancia por Permutación (MDA). La primera mide la reducción promedio de impureza (índice de Gini) que aporta una variable durante el entrenamiento, mientras que la segunda evalúa la disminución del rendimiento del modelo al alterar

aleatoriamente los valores de una variable (Strobl, Boulesteix, Zeileis, y Hothorn, 2007). Aunque la Importancia por Permutación es más robusta, resulta computacionalmente más costosa (Breiman, 2001); por ello, en este estudio exploratorio se emplea la Importancia Gini, reconocida por su eficiencia y disponibilidad en herramientas como scikit-learn, dejando abierta la posibilidad de futuras validaciones con MDA para fortalecer la robustez de los resultados.

## Indicadores Financieros para el Análisis Sectorial

Los indicadores financieros son cálculos que comparan cifras de los estados financieros para medir el rendimiento y la salud financiera (Ross, Westerfield, y Jaffe, 2016). Para este estudio se seleccionaron ratios representativos agrupados en cuatro categorías:

**(a) Liquidez:** que mide la capacidad de la empresa para cubrir obligaciones a corto plazo mediante la Razón Corriente, la Prueba Ácida, el Ratio de Efectivo y el Capital de Trabajo.

**(b) Rentabilidad:** que evalúa la generación de utilidades a través del Retorno sobre Activos (ROA), Retorno sobre el Patrimonio (ROE), Margen Bruto (MB), Margen Operativo (MO) y el Margen de Utilidad Neta.

**(c) Eficiencia:** que analiza el uso de los activos para generar ventas mediante la Rotación de Inventarios, Rotación de Cuentas por Cobrar, Rotación de Activos.

**(d) Solvencia:** que refleja la estructura financiera a través del Endeudamiento Total, Autonomía financiera, Razón de Endeudamiento y la Cobertura de Interés. Estos indicadores proporcionan una visión integral del desempeño financiero y sirven como variables predictoras clave dentro del modelo Random Forest.

## Metodología

### Fuente y Selección de Datos

La **fuente** principal de datos para este estudio son los estados financieros trimestrales que las empresas del sector minero presentan públicamente a la Superintendencia del Mercado de Valores (SMV) del Perú. La **muestra** de estudio comprende a 14 empresas clasificadas en el sector minero que cuentan con información financiera completa y auditada para el período que abarca desde el primer trimestre de 2020 hasta el último trimestre de 2024.

### Definición de Variables Operacionales

#### Variables Predictoras (X)

Corresponden al conjunto de indicadores financieros calculados. Cada observación en el conjunto de datos representa una empresa en un período de tiempo específico y las columnas corresponden a los valores calculados de los ratios de liquidez, rentabilidad, eficiencia y solvencia.

#### Variable Objetivo (Y)

Para construir un modelo predictivo, se operacionalizó el concepto de "Riesgo Financiero Futuro" como una variable categórica binaria. La variable se definió de la siguiente manera: para cada observación (empresa/trimestre), se identificó el Retorno sobre Activos (ROA) del siguiente trimestre (t+1). Se estableció un umbral de rentabilidad mínima de 5%. El desempeño se clasificó como: **1 (Riesgo / Mal Desempeño):** Si su ROA\_Futuro fue **inferior** a 5%. Y **0 (Sin Riesgo / Buen Desempeño):** Si su ROA\_Futuro fue **superior** a 5%.

### Preprocesamiento de Datos

#### Construcción y Entrenamiento del Modelo

El flujo de trabajo para la implementación del modelo fue el siguiente:

División de Datos: el conjunto de datos preprocesado se dividió en un conjunto de entrenamiento (70% de los datos) y un conjunto de prueba (30% restante).

Implementación del Algoritmo: se utilizó la clase RandomForestClassifier de la biblioteca de machine learning scikit-learn en Python. Se configuraron los hiperparámetros clave del modelo, incluyendo n\_estimators=100 (el número de árboles en el bosque) y random\_state para garantizar la reproducibilidad de los resultados.

Entrenamiento: el modelo fue entrenado invocando el método fit() sobre el conjunto de datos de entrenamiento (X\_train, y\_train).

Evaluación del Desempeño del Modelo: el desempeño del modelo se evaluó con el conjunto de prueba (30% de los datos) utilizando la métrica de **precisión (accuracy)**, que mide la proporción de observaciones correctamente clasificadas. También se revisó la **matriz de confusión** para verificar el nivel de acierto en las categorías "Riesgo" y "Sin Riesgo".

Extracción de Importancia: una vez finalizado el entrenamiento, se extrajeron los puntajes de importancia Gini para cada indicador financiero utilizando el atributo "desempeño\_target" del objeto del modelo ya ajustado.

## Resultados

### Efectividad Predictiva del Modelo

La Figura 1 presenta el rendimiento del Random Forest, evaluados sobre el conjunto de prueba.

Figura N° 1: Métricas de rendimiento del modelo de Random Forest.

| --- Evaluación del Modelo (Datos de Prueba) --- |           |        |          |         |
|---|-----------|--------|----------|---------|
|   | precision | recall | f1-score | support |
| 0   | 0.83      | 0.96   | 0.89     | 52      |
| 1   | 0.90      | 0.64   | 0.75     | 28      |
| accuracy  |           |        | 0.85     | 80      |
| macro avg                                       |           | 0.87   | 0.80     | 80      |
| weighted avg                                    |           | 0.86   | 0.85     | 80      |

Fuente: Ilustración obtenida de los resultados en Google Collaboratory. A continuación, se describe textualmente la estructura que una figura gráfica representaría.

El modelo obtuvo una precisión del 85%, lo que refleja un buen desempeño predictivo. Mostró mayor capacidad para identificar empresas con buen desempeño (recall = 0.96) y una precisión aceptable en los casos de riesgo (recall = 0.64), confirmando la eficacia general del algoritmo Random Forest.

Tabla N°1: Ranking de Importancia de Indicadores Financieros.

| Rango | Categoría del Indicador | Indicador Financiero           | Importancia Gini Normalizada |
|-------|-------------------------|--------------------------------|------------------------------|
| 1     | Rentabilidad            | Retorno sobre activos (ROA)    | 0.1323                       |
| 2     | Rentabilidad            | Margen operativo (MO)          | 0.1302                       |
| 3     | Rentabilidad            | Margen neto                    | 0.1291                       |
| 4     | Rentabilidad            | Margen bruto (MB)              | 0.0898                       |
| 5     | Rentabilidad            | Rentabilidad sobre patrimonio  | 0.0683                       |
| 6     | Eficiencia              | Rotación de cuentas por cobrar | 0.0631                       |
| 7     | Liquidez                | Razón de efectivo              | 0.0606                       |
| 8     | Eficiencia              | Rotación de Inventarios        | 0.0544                       |
| 9     | Solvencia               | Cobertura de Intereses         | 0.0471                       |
| 10    | Liquidez                | Razón corriente                | 0.0438                       |
| 11    | Liquidez                | Prueba ácida                   | 0.0385                       |
| 12    | Solvencia               | Endeudamiento total            | 0.0331                       |
| 13    | Solvencia               | Autonomía financiera           | 0.0321                       |
| 14    | Eficiencia              | Rotación de activos            | 0.0312                       |
| 15    | Solvencia               | Razón de endeudamiento         | 0.0257                       |
| 16    | Liquidez                | Capital neto                   | 0.0207                       |

Fuente: Elaboración propia obtenida de los resultados de los modelos Árbol de Decisión y Random Forest, en Google Collaboratory. A continuación, se describe el cuadro que ordena las variables de mayor a menor importancia.

### Relevancia de los Ratios según Random Forest

La Tabla 1 muestra el ranking de los indicadores financieros según su capacidad predictiva, ordenados de mayor a menor según su puntaje de Importancia Gini normalizado.

El modelo Random Forest evidenció que los **ratios de rentabilidad** son los factores más influyentes para predecir el desempeño futuro del sector minero peruano. Indicadores como el **ROA, ROE, Margen Neto, Margen Operativo y Margen Bruto** demostraron una fuerte **persistencia temporal**, lo que indica que la rentabilidad actual de una empresa es la mejor señal de su rentabilidad en el corto plazo. Este comportamiento refleja la continuidad del desempeño financiero entre períodos consecutivos.

En un segundo nivel de importancia, se ubicaron los **ratios de eficiencia**, como la Rotación de Cuentas por Cobrar y la Rotación de Activos, los cuales aportan información adicional sobre el uso eficiente de los recursos. En cambio, los indicadores de **liquidez y endeudamiento** mostraron menor poder predictivo, debido a la estabilidad financiera característica de las grandes empresas mineras, lo que limita su capacidad para anticipar variaciones significativas en la rentabilidad. Por lo tanto, estos ratios no ofrecen la varianza o la señal de "alerta temprana" que sí ofrecen los indicadores de rentabilidad, que son más dinámicos y están directamente ligados al desempeño operativo que se busca predecir.

### Conclusión

El modelo predictivo basado en Random Forest permitió identificar los indicadores financieros con mayor poder explicativo sobre el desempeño futuro del sector minero peruano. Los resultados evidencian que los ratios de rentabilidad (ROA, ROE, Margen Neto, Margen Operativo y Margen Bruto) son las variables más influyentes para anticipar el riesgo de bajo desempeño (ROA < 5%), lo que confirma la persistencia temporal de la rentabilidad en el corto plazo.

En contraste, los indicadores de liquidez y endeudamiento mostraron una influencia limitada, debido a la estabilidad financiera de las grandes empresas del sector, mientras que los ratios de eficiencia ofrecieron información complementaria sobre la gestión operativa. El modelo alcanzó una precisión del 85%, validando su utilidad para tareas de predicción financiera.

Estos hallazgos aportan una herramienta útil para la evaluación de riesgos y toma de decisiones en la industria minera, al permitir identificar tempranamente señales de deterioro del desempeño financiero. Futuras investigaciones podrían incorporar métodos de validación cruzada, comparar otros algoritmos de Machine Learning o ampliar la muestra a otros sectores productivos para fortalecer la generalización de los resultados.

## Bibliografía

BBVA Research. (3 de febrero de 2023). Perú Situación de sector minero-2022. <https://www.bbvareresearch.com/wp-content/uploads/2023/02/Situacion-del-sector-minero-2022-1.pdf>

Breiman, L. (2001). Random forest. Machine Learning, 45(1), pp. 5-32. <https://doi.org/10.1023/A:1010933404324>

Goodfellow, I., Bengio, Y. y Courville, A. (2018). Deep learning. Genet Program Evolvable Mach, 19, pp. 305-307. <https://doi.org/10.1007/s10710-017-9314-z>

Hastie, T., Tibshirani, R. y Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed. ed.). Springer.

James, G., Witten, D., Hastie, T. y Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>

Lizarzaburu, E., Noriega, E., Macias, W., Rodríguez, K., y Berggrun, L. (2019). Innovación y finanzas en empresas mineras de latinoamérica. Revista Científica Electrónica de Ciencias Humanas, 50-70. [www.revistaorbis.org.ve](http://www.revistaorbis.org.ve)

Ministerio de Energía y Minas. (24 de mayo de 2018). Minería peruana representó alrededor del 10% del PBI y cerca del 62% del valor total de las exportaciones en el 2017. <https://www.gob.pe/institucion/minem/noticias/1362-mineria-peruana-represento-alrededor-del-10-del-pbi-y-cerca-del-62-del-valor-total-de-las-exportaciones-en-el-2017>

Ross, S., Westerfield, R. y Jaffe, J. (2016). Corporate Finance (11th ed.). McGraw-Hill Education.

Strobl, C., Boulesteix, A., Zeileis, A. y Hothorn, T. (2007). Bias in random forest variable importance measures: We need a conditional approach. *BCM Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>

Sun, G. y Deng, S. (2025). Financial Time Series Forecasting: A Comparison Between Traditional Methods and AI-Driven Techniques. *Journal of Computer Signal and System Research*, 2(2), pp. 86-93. <https://doi.org/10.71222/339b9812>

Velasquez, J. L., Zegarra, C., y Fernandez, J. (2021). Mine 2021: Grandes expectativas, aprovechando el mañana. PwC Perú