



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Estado de México

**Análisis y Reporte sobre el desempeño del modelo**

TC3006C. Inteligencia artificial avanzada para la ciencia de datos I

Grupo: 101

A01749667. Alan Contreras Prieto

Prof. Jorge Adolfo Ramírez Uresti

Fecha de entrega: 10 de septiembre de 2025

Semestre agosto – diciembre 2025

# **Análisis del desempeño del modelo**

## **Separación de los datos y evaluación del modelo**

Para evaluar el desempeño del modelo, el dataset se dividió en dos conjuntos: 80% para entrenamiento: con el fin de que el modelo pueda aprender patrones y relaciones dentro de los datos. 20% para prueba: utilizado exclusivamente para validar el rendimiento y asegurar que el modelo no se limite a memorizar los ejemplos de entrenamiento. Esta proporción es una práctica común en problemas de aprendizaje automático, pues ofrece un equilibrio entre disponer de un volumen amplio de datos para entrenar y conservar una muestra representativa para medir la capacidad de generalización del modelo.

La evaluación del desempeño se llevó a cabo mediante el uso de la matriz de confusión y de métricas fundamentales de clasificación, tales como accuracy, precision, recall y F1-score, obteniendo como resultado un accuracy general de 0.83, lo que refleja un rendimiento sólido y confiable.

## **Diagnóstico de bias (sesgo)**

En cuanto al diagnóstico de bias o sesgo, este concepto hace referencia a los errores que aparecen cuando un modelo realiza suposiciones demasiado simples o cae en generalizaciones excesivas, limitando su capacidad predictiva. En este caso, el uso de un Random Forest resulta especialmente adecuado, ya que este modelo combina múltiples árboles de decisión y, al hacerlo, logra reducir de manera significativa el sesgo característico que suele presentarse en un único árbol.

Por esta razón, el bias se puede diagnosticar como bajo, lo que se evidencia en el nivel de accuracy obtenido y en la distribución equilibrada de resultados entre las clases. Estos indicadores sugieren que el modelo fue capaz de captar las relaciones presentes en los datos sin recurrir a simplificaciones inadecuadas.

## **Diagnóstico de varianza**

Respecto a la varianza, esta se entiende como la sensibilidad del modelo a pequeñas variaciones en los datos de entrenamiento, lo cual, si es elevado, puede conducir al sobreajuste. En el caso de Random Forest, la construcción del modelo a partir de un ensamble de árboles entrenados con subconjuntos aleatorios de datos y características contribuye a mitigar este riesgo, ya que el promedio de las predicciones reduce la inestabilidad inherente a cada árbol individual.

Por esta razón, el nivel de varianza se puede diagnosticar como medio-bajo. Aunque los árboles por separado tienden a mostrar alta varianza, el efecto de agregación característico del Random Forest permite suavizar este comportamiento. El desempeño estable en el conjunto de prueba respalda esta conclusión, confirmando que el modelo no depende en exceso de las particularidades del conjunto de entrenamiento.

## **Diagnóstico del ajuste del modelo**

El ajuste del modelo se analiza comparando los resultados en entrenamiento y prueba:

- Underfitting: ocurre cuando el modelo no logra aprender los patrones de los datos y obtiene un bajo rendimiento tanto en entrenamiento como en prueba.
- Overfitting: ocurre cuando el modelo memoriza los datos de entrenamiento y falla al generalizar, mostrando una gran diferencia de desempeño entre entrenamiento y prueba.
- Fit adecuado: ocurre cuando el modelo logra un equilibrio entre aprendizaje y generalización.

Por lo tanto, el ajuste del modelo se puede diagnosticar como un Fit adecuado. El modelo alcanzó un desempeño consistente y sólido entre el conjunto de entrenamiento y el de prueba, sin diferencias significativas. Los resultados son equilibrados entre todas las clases, lo que confirma que el Random Forest fue capaz de generalizar sin caer en underfitting ni en overfitting.

## Técnicas de regularización o ajuste de parámetros

Para mejorar el desempeño del modelo se aplicó un proceso de ajuste de hiperparámetros utilizando la técnica de GridSearchCV, la cual permite evaluar de manera sistemática distintas combinaciones de parámetros para identificar aquellas que generan un mejor rendimiento. En este caso, se definió una cuadrícula de búsqueda que incluyó variaciones en parámetros clave del modelo Random Forest:

- El número de estimadores (n\_estimators) [10, 50, 100]: Un valor pequeño, como 10, genera un modelo más ligero y rápido, pero con menor capacidad de generalización porque depende mucho de cada árbol individual. A medida que se incrementa este valor (50, 100), el modelo tiende a ser más estable y preciso, ya que el promedio de más árboles reduce la varianza de las predicciones. Sin embargo, esto también aumenta el costo computacional, por lo que resulta importante encontrar un equilibrio entre rendimiento y tiempo de entrenamiento. En este caso, se redujo el rango de búsqueda a un máximo de 100 árboles porque con valores más altos el tiempo de ejecución de GridSearch se volvía excesivo.
- El criterio de división (criterion) ['gini', 'entropy', 'log\_loss']: El criterio de Gini es el más eficiente computacionalmente y suele funcionar bien en la práctica al medir la impureza de los nodos. La entropía, basada en la teoría de la información, produce divisiones más equilibradas pero con un mayor costo de cálculo. Finalmente, log\_loss (o devianza) está más relacionado con la optimización de la probabilidad de clasificación y resulta útil en problemas multiclase, como el que se abordó en esta práctica. Evaluar los tres criterios permite comparar el rendimiento e impacto en el tiempo de entrenamiento.
- La profundidad máxima de los árboles (max\_depth) [10,20,30, None]: Profundidades pequeñas, como 10, limitan la complejidad del modelo y ayudan a reducir el riesgo de sobreajuste, pero podrían perder información relevante en datasets más complejos. Por otro lado, valores intermedios como 20 o 30 permiten capturar relaciones más profundas entre las variables sin llegar a memorizar los datos. Finalmente, dejar la profundidad sin restricción (None) permite que cada

árbol crezca hasta su máximo, lo cual puede mejorar el ajuste en entrenamiento pero con el riesgo de incrementar la varianza y sobreajustar los datos.

- La cantidad mínima de muestras requeridas para dividir un nodo (`min_samples_split`) [50,100,250,500]: Con valores bajos, los árboles tienden a crecer más y ajustarse demasiado a los datos. Con valores más altos, como 250 o 500, se fuerza al modelo a crear divisiones solo cuando hay suficiente información, lo que reduce el sobreajuste y mejora la capacidad de generalización. En este caso, incluir valores desde 50 hasta 500 permitió evaluar un rango balanceado entre flexibilidad y regularización.

Esta exploración exhaustiva permitió comparar arquitecturas distintas y también encontrar un balance adecuado entre complejidad del modelo y capacidad de generalización.

Tras la ejecución de `GridSearchCV` se seleccionó el conjunto de hiperparámetros que ofreció la mejor precisión promedio en la validación cruzada, lo cual se tradujo en una mejora del desempeño del modelo en el conjunto de prueba.