



Tecnológico de Monterrey

Momento de Retroalimentación: Módulo 2

Análisis y Reporte del modelo Random Forest

Materia:

Inteligencia artificial avanzada para la ciencia de datos I
(Gpo 101)

Alumno:

Alfredo Azamar López - A01798100

Profesor:

Jorge Adolfo Ramírez Uresti

Dataset

Nota:

Para la entrega pasada el código utiliza un dataset relacionado al tema de cáncer de mama (la base de datos se extrae de una librería de *sklearn*), por motivos del análisis del desempeño se cambia el dataset al de “Spaceship Titanic” (obtenido de [Kaggle](#))

Este nuevo dataset tiene información relacionada a pasajeros en un entorno espacial ficticio, el cual incluye datos como su planeta destino, edad, gastos en varios servicios, si son clientes VIP, etc. La variable a predecir es si el pasajero fue transportado o no, esto lo convierte en un problema de clasificación binaria, el cual se adapta perfectamente a nuestro modelo de Random Forest debido a que crean múltiples árboles de decisión y se utiliza el voto mayoritario de estos para tomar una decisión final obteniendo un resultado robusto y preciso.

Las propiedades de los valores dentro del dataset son diversas ya que se incluyen variables numéricas como categóricas que influyen en la clasificación, Random Forest aprovecha esta característica y se adapta a la combinación de estas variables, siendo capaz de capturar las relaciones complejas entre las variables sin necesidad de que se especifiquen explícitamente las interacciones.

Separación y evaluación del modelo

Para este problema el dataset está dividido en un conjunto de entrenamiento y prueba en una relación de 80/20, para la evaluación del modelo se utilizaron las métricas de exactitud, reporte de clasificación, matriz de confusión y curvas de aprendizajes.

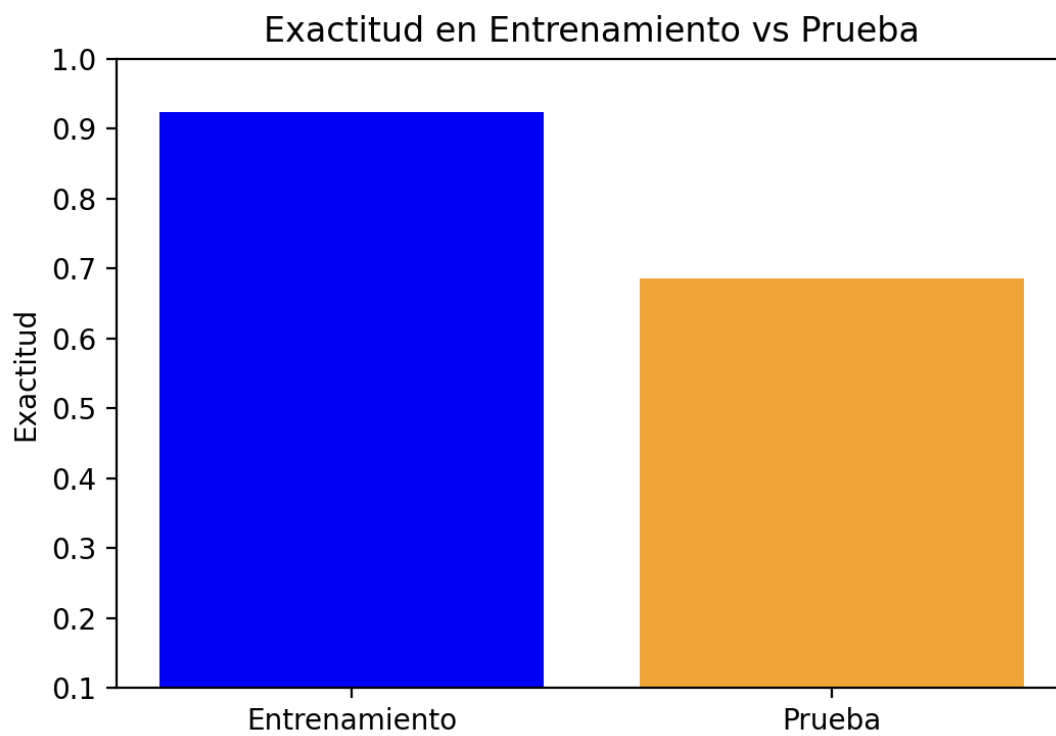


Imagen 1. Gráfica sobre la exactitud del dataset (dividido en entrenamiento y prueba)

Diagnóstico del grado de Bias

Diagnóstico del grado de Varianza

Diagnóstico del nivel de ajuste del modelo

Técnicas de regularización