

# Analisis de Resultados del Proyecto: Rendimiento Académico de Estudiantes

---

## Objetivo General

Desarrollar un modelo de Aprendizaje Automático capaz de predecir si un estudiante aprobará o no (variable binaria: **Pass / Fail**) utilizando variables académicas. Esto permitirá identificar estudiantes en riesgo y actuar de forma preventiva.

## 1. Comparación de Modelos de Clasificación

Se evaluaron tres modelos de aprendizaje automático:

Modelo	Precisión (Accuracy)	Observaciones
Regresión Logística	0.49 (49%)	Bajo desempeño; no logra distinguir adecuadamente entre estudiantes que aprueban y reprueban. Métricas equilibradas pero poco informativas.
Árbol de Decisión	0.49 (49%)	Mismo nivel de precisión general, pero con <b>recall muy alto para “Reprobar” (0.90)</b> . Esto indica que identifica bien a los estudiantes que están en riesgo, aunque falla mucho con quienes sí aprueban.
Random Forest	0.66 (66%)	Fue el modelo con <b>mejor desempeño global</b> . Presentó un buen <b>equilibrio entre precisión, recall y F1-score</b> , tanto para “Aprobar” como “Reprobar”. Se destaca por su robustez frente a variaciones en los datos.

**Conclusión parcial:** El modelo de **Random Forest** fue el más efectivo, tanto en precisión general como en balance entre clases. Por esta razón se lo eligió como modelo final para realizar predicciones.

## 2. Análisis de Perfiles de Estudiantes que Reprueban

El análisis de los estudiantes etiquetados como “Desaprueba” (`final_result = 0`) mostró un perfil bastante definido:

### Tendencias observadas:

- **Menores puntajes académicos** en matemáticas, lectura y escritura (promedios alrededor de 74 sobre 100).
- **Asistencia muy baja**, con valores anormalmente altos por errores de escala (e.g.,  $7.89e+13$ ), lo que sugiere una **necesidad de corrección o normalización** más adecuada.
- **Horas de estudio también bajas**, con problemas similares de escala.

### Visualización con boxplots:

- Se evidenció que los estudiantes que reprueban tienen claramente **menor asistencia** y **menos horas de estudio** que los que aprueban.
- Esta información respalda la hipótesis de que el compromiso con el estudio y la participación regular en clase son factores críticos.

**Conclusión parcial:** El modelo permite identificar patrones de riesgo. La asistencia y las horas de estudio, cuando están por debajo del promedio, se relacionan directamente con mayor probabilidad de desaprobación.

---

### 3. Importancia de las Variables Predictoras (Random Forest)

El modelo Random Forest también permitió determinar **cuáles variables fueron más útiles para predecir el rendimiento académico**.

Esto revela algunos puntos clave:

- Aunque la **asistencia** y las **horas de estudio** son importantes, **los puntajes académicos tienen mayor peso** en la predicción, especialmente los de **lectura** y **escritura**.
- El puntaje de **lectura** fue el predictor más fuerte. Esto sugiere que una buena comprensión lectora puede ser un factor base que afecta el rendimiento en varias materias.

**Conclusión parcial:** Las variables académicas (notas) influyen más que los hábitos (asistencia/estudio), aunque estos también son relevantes. Lectura y escritura se consolidan como los factores más influyentes.

### Conclusión General del Análisis

- El modelo **Random Forest** logró una **predicción razonable (66%)**, lo que lo convierte en una herramienta útil para detectar estudiantes en riesgo.
- El perfil de los estudiantes que reprueban es claro: **baja asistencia, pocas horas de estudio y bajos puntajes**.
- Las variables más influyentes fueron los **puntajes en lectura, escritura y matemáticas**, pero la **asistencia** también puede marcar la diferencia.
- Este análisis confirma que **es posible anticipar el fracaso académico utilizando modelos de datos**, lo que abre la puerta a estrategias de intervención temprana.