https://snies.mineducacion.gov.co/portal/ESTADISTICAS/Bases-consolidadas/

GÉNERO, MATRICULADOS

Género

One-Hot (One-Hot Encoding)

737.690 datos

¡Excelente pregunta! Analicemos los indicadores de confiabilidad de tu modelo:

* **Error Absoluto Medio (MAE)**: Este valor (8.46 estudiantes) representa la diferencia promedio absoluta entre el número real de estudiantes matriculados y el número que tu modelo predijo. En promedio, las predicciones del modelo se desvían en aproximadamente 8.46 estudiantes de los valores reales. Un MAE más bajo indica un mejor rendimiento.
* **Error Cuadrático Medio (MSE)**: Este valor (183.12) mide el promedio de los errores al cuadrado. Penaliza más los errores grandes. Es útil para entender la varianza de los errores.
* **Raíz del Error Cuadrático Medio (RMSE)**: Este valor (13.53 estudiantes) es la raíz cuadrada del MSE. Está en las mismas unidades que la variable objetivo (número de estudiantes), lo que lo hace más interpretable que el MSE. Indica la dispersión típica de los errores de predicción.
* **R-squared (R²)**: Este valor (0.99) es un indicador de qué tan bien se ajusta el modelo a los datos. Un valor de R² cercano a 1 (como en este caso) sugiere que el modelo explica una gran proporción de la variabilidad en el número de estudiantes matriculados. Un R² de 0.99 es muy alto y generalmente indica un excelente ajuste del modelo a los datos de entrenamiento y prueba.

**¿Indican un buen modelo?**

Sí, basándonos en estos indicadores, especialmente en el alto valor de R² (0.99) y un MAE relativamente bajo en comparación con el rango de tus datos, el modelo parece tener un **muy buen rendimiento** para predecir el número total de matriculados. Explica casi el 99% de la variabilidad en los datos, lo cual es excelente.

Sin embargo, siempre es importante considerar el contexto del problema y los datos. Aunque las métricas son muy buenas, es recomendable:

* Visualizar las predicciones frente a los valores reales para identificar posibles patrones de error.
* Considerar si un MAE de 8.46 estudiantes es aceptable para tu aplicación específica.

En resumen, los datos de la tabla sugieren que tu modelo de Random Forest es bastante confiable para esta tarea de predicción.

¿Te gustaría visualizar las predicciones del modelo frente a los valores reales para tener una idea más clara de su rendimiento?