

# MIAD



Maestría  
en Inteligencia  
Analítica de Datos

## **Despliegue de soluciones analíticas**

# **Análisis de los resultados de Evaluar para Avanzar 2022-1 para los estudiantes de instituciones educativas de los departamentos de la Amazonía colombiana**

**Grupo 11**  
**Proyecto final**

**Dayron Cuadros, Fabricio Morales, Daniel Londoño y David Ruiz**

# Problema de negocio

Evaluar para Avanzar en la Amazonía Colombiana



# Evaluar para Avanzar

El **Icfes** ha desarrollado una estrategia complementaria a las Pruebas Saber conocida como **Evaluar para Avanzar**. La implementación de Evaluar para Avanzar se realiza entre los grados 3° y 11°. En estos grados se aplican instrumentos de valoración asociados a las **áreas básicas de conocimiento**.

Instrumento de valoración	Grados								
	3°	4°	5°	6°	7°	8°	9°	10°	11°
Competencias Comunicativas en Lenguaje: Lectura									
Matemáticas									
Ciencias Naturales y Educación Ambiental									
Competencias Ciudadanas: Pensamiento Ciudadano									
Inglés									
Cuestionarios Auxiliares									

## Resultados de Evaluar para Avanzar

El problema de negocio consiste en analizar los resultados de los distintos instrumentos de los **estudiantes** de las **instituciones educativas** de los departamentos que componen la Amazonía colombiana (**Guainía, Vaupés, Vichada, Guaviare y Amazonas**) que aplicaron la estrategia de Evaluar para Avanzar en el **2022-1**.

## Predicción del desempeño en Lectura Crítica

Además, se implementaron modelos de *machine learning* para intentar **predecir los resultados de la prueba de Lectura Crítica** a partir de una caracterización del estudiante, considerando que en algunas instituciones no siempre hay disponibilidad de tiempo y recursos para aplicar todos los instrumentos a todos los estudiantes.

# Modelos desarrollados

Evaluar para Avanzar en la Amazonía Colombiana

## Primeros modelos

El equipo consideró los siguientes **algoritmos** de manera inicial para desarrollar los **modelos de predicción** del desempeño en **Lectura Crítica**:

- Modelo de regresión lineal múltiple
- Modelo de red neuronal
- Modelo de bosque aleatorio (*Random Forest*)
- Modelo de árbol de decisión (*Decision Tree*)
- Modelo de soporte vectorial (*Support Vector Machines*)



## Evaluación del desempeño de los modelos

Para evaluar el desempeño de los modelos considerados bajo los distintos algoritmos y escenarios se eligieron dos métricas: El **Score** y el **MSE**.

A partir de estas métricas, se identificó que los modelos de **bosque aleatorio** (*Random Forest*) eran los de mejor desempeño entre las opciones consideradas.

## Refinamiento de los modelos

A partir de los resultados obtenidos en los primeros modelos, el equipo decidió implementar tres **algoritmos basados en árboles de decisión** debido a su capacidad para reconocer patrones complejos:

- Modelo de *Random Forest*
- Modelo de *Gradient Boosting*
- Modelo de *XGBoost*

## Selección del modelo final

Al evaluar los modelos desarrollados usando las **métricas** definidas previamente, se obtuvieron los siguientes resultados:

Algoritmo	Score	MSE
Random Forest	0,603	0,440
Gradient Boosting	0,635	0,383
XGBoost	0,639	0,385

## Selección del modelo final

En conclusión, se observa que el **algoritmo de XGBoost** presenta un mejor desempeño. Por esta razón, se decidió aplicar **GridSearchCV** para encontrar los **parámetros óptimos**, que fueron los siguientes:

- ***n\_estimators***: 200
- ***learning\_rate***: 0.2
- ***max\_depth***: 4



# Tablero de control

Evaluar para Avanzar en la Amazonía Colombiana

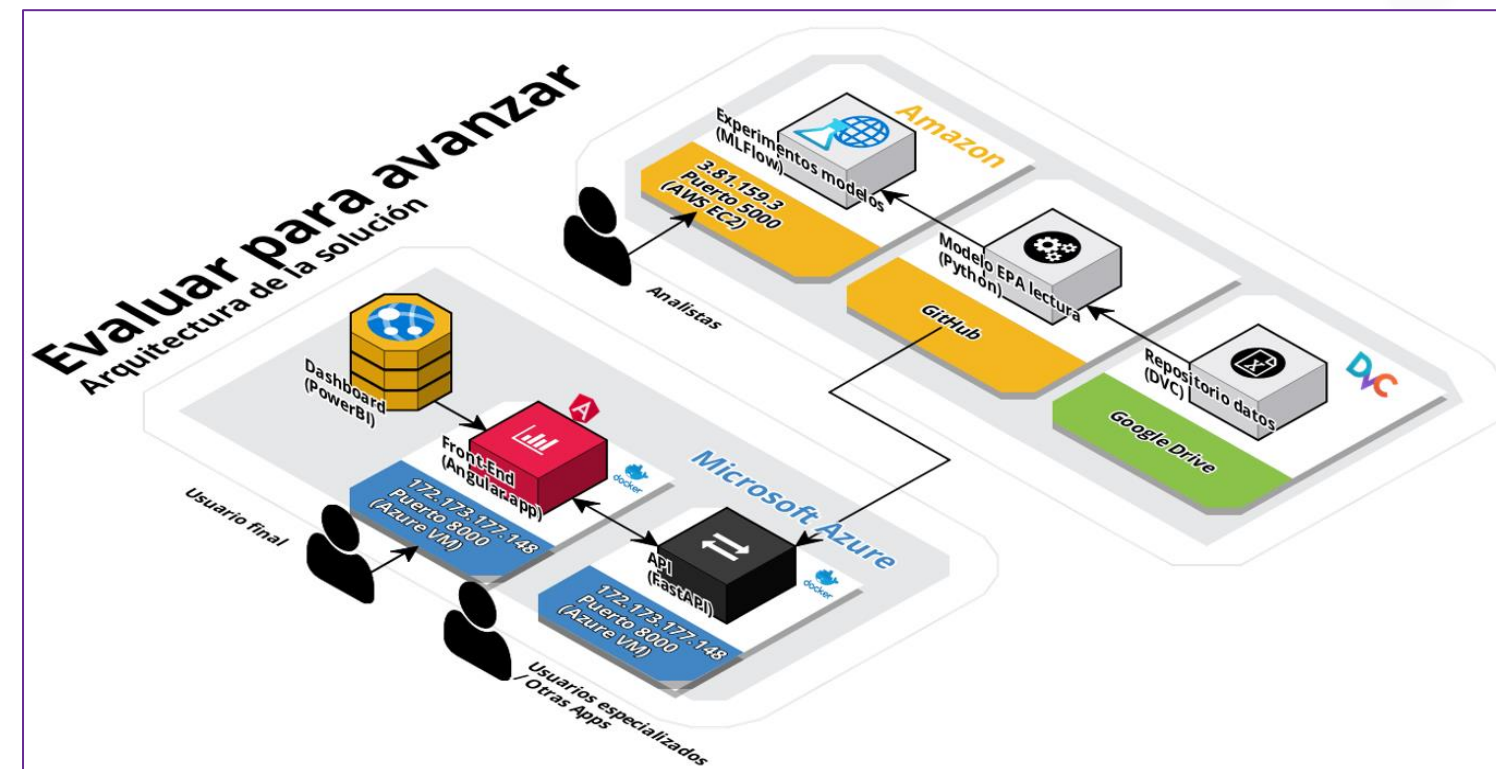
## Despliegue del tablero

Para el despliegue de la solución se optó por la estrategia de **contenedores Docker** sobre **Infraestructura como Servicio** ofrecida por Azure.

Para el **entrenamiento de modelos** se utilizó un servicio de **IaaS** en Amazon, particularmente el servicio **EC2** para publicar un servidor **MLFlow** a través del cual se realizaron experimentos.

# Despliegue del tablero

Esta decisión de arquitectura de la solución estuvo alineada con la necesidad de mantener estos dos componentes disponibles para los usuarios finales el mayor tiempo posible.



## **Análisis exploratorio de los resultados**

---

La primera sección del tablero incluye información los estudiantes evaluados, y los resultados de los instrumentos de competencias ciudadanas, ciencias naturales, matemáticas y nuestro foco: lectura crítica.

La herramienta incluye gráficos y tablas que permiten visualizar los datos de forma clara y concisa.



## Predicción del desempeño en Lectura Crítica

En la parte inferior derecha se puede acceder a la funcionalidad de predicción, en la que, a partir del grado del estudiante, el género, su institución educativa y el número de respuestas correctas en los otros tres instrumentos se puede acceder a una predicción sobre el instrumento de Lectura Crítica: Bajo, Medio o Alto.

# Resultados y conclusiones

Evaluar para Avanzar en la Amazonía Colombiana

## Modelos desarrollados

- Se probaron varios conjuntos de variables y algoritmos para desarrollar modelos de predicción para el desempeño de los estudiantes en Lectura Crítica. Finalmente, se implementó un modelo utilizando el algoritmo XGBoost.

## Información descriptiva

- Se creó un panel interactivo que ofrece información relevante sobre los resultados de los estudiantes de las instituciones educativas en los departamentos de la Amazonía colombiana.



## Futuros trabajos

- Existe la posibilidad de ampliar las funciones del tablero para incluir modelos para predicciones de Matemáticas, Ciencias Naturales y de Competencias Ciudadanas. Así mismo, se puede considerar la incorporación del análisis para los 32 departamentos de Colombia.



**MIAD**



Maestría  
en Inteligencia  
Analítica de Datos

