

## ## Contexto

Problema: Comprender y modelar modalidad de programas y edad promedio, explorando también estructura

Fecha de generación: 2025-11-21T03:08:32.998373Z

Leakage: Leakage flag=False r2=0.18854441022979995 features=['PROMEDIO EDAD HOMBRE ', 'PRO

## ## Dataset

Matriz de correlación: 200 variables

Fuente: data/raw/\*.csv (ver config/config.py para parámetros). Tamaño aproximado post-proceso: ver X\_

## ## Metodología

Pipeline: EDA -> Preprocesamiento -> Feature Engineering -> Entrenamiento -> Evaluación -> Interpretación

Validación temporal aplicada (split por año). HPO: HPO ejecutado (ver hpo\_summary.md para top config

## ## Resultados

### ### Clasificación

Métrica: AUC-ROC=0.4644

Figura/Importancias: ver reports/feature\_importance\_classification.csv (top 5 abajo si disponible).

### ### Regresión

Métrica: pendiente

Figura/Importancias: ver reports/feature\_importance\_regression.csv.

### ### Clustering

Top clustering (silhouette):

algo|params|silhouette|homogeneity|completeness|runtime\_ms

DBSCAN|{'eps': 0.7, 'min\_samples': 5}|0.2192823255920242|||737.24

KMeans|{'n\_clusters': 7}|0.1822504906134569|||237.89

DBSCAN|{'eps': 0.7, 'min\_samples': 10}|0.1820721923050643|||753.8

Figura: clustering\_dbscan\_eps\_grid.png

### ### Anomalías

Anomalías (fracción vs target):

algo|anomaly\_fraction|contamination\_cfg|runtime\_ms

IsolationForest|0.05|0.05|1033.32

LocalOutlierFactor|0.05|0.05|851.72

LocalOutlierFactor|0.05|0.05|869.44

EllipticEnvelope|0.05|0.05|1813.7

Figura: anomaly\_fraction\_by\_algo.png

## ## Interpretabilidad

Top importancias clasificación:

feature|importance

CARRERA CLASIFICACIÓN NIVEL 2\_Carreras Técnicas|0.2572521855202013

CLASIFICACIÓN INSTITUCIÓN NIVEL 1\_Universidades|0.2394819847710022

CARRERA CLASIFICACIÓN NIVEL 1\_Técnico de Nivel Superior|0.1896091493932357

NOMBRE INSTITUCIÓN\_IP AIEP|0.1038523508856253

CLASIFICACIÓN INSTITUCIÓN NIVEL 2\_Universidades CRUCH|0.0632917366810482

Importancias regresión: pendiente

## ## Recomendaciones

- Consolidar almacenamiento de modelos entrenados y versionar.
- Incorporar monitoreo de drift y recalibración anual.

- Extender modelos a boosting y SHAP para interpretabilidad avanzada.
- Integrar panel UI para exploración de anomalías y clusters.