

Aplicación de Modelo de Regresión y Optimización con Optuna en Datos de COVID-19 del Perú

Universidad Nacional del Altiplano
Curso: Métodos de Optimización
Estudiante: Kenny Leonel Ccora Quispe

May 28, 2025

1 Introducción

La predicción de casos diarios de COVID-19 se ha convertido en una tarea crítica para la toma de decisiones en salud pública. En este estudio, se utilizan modelos de regresión aplicados a un conjunto de datos abiertos del Ministerio de Salud del Perú, complementando con optimización de hiperparámetros mediante `Optuna`.

2 Dataset y Preparación

Se utilizó el dataset **Casos positivos por COVID-19** del MINSA. Este contiene información por persona: sexo, edad, distrito y fecha del resultado positivo. Se agrupó por distrito y fecha para generar la variable objetivo **casos diarios**. Además, se incluyeron variables como edad promedio, proporción de hombres y día de la semana.

3 Metodología

3.1 Regresión Aplicada

Se probaron distintos modelos de regresión para predecir los casos diarios a partir de las variables generadas.

3.2 Comparación de Modelos

Se comparó el rendimiento con **Random Forest**. El segundo fue optimizado con `Optuna`. La métrica principal usada fue el **coeficiente de determinación (R^2)**.

3.3 Optimización con Optuna

Optuna se utilizó para explorar hiperparámetros de Random Forest, incluyendo `n_estimators`, `max_depth`, y `max_features`, obteniendo mejoras significativas.

4 Resultados

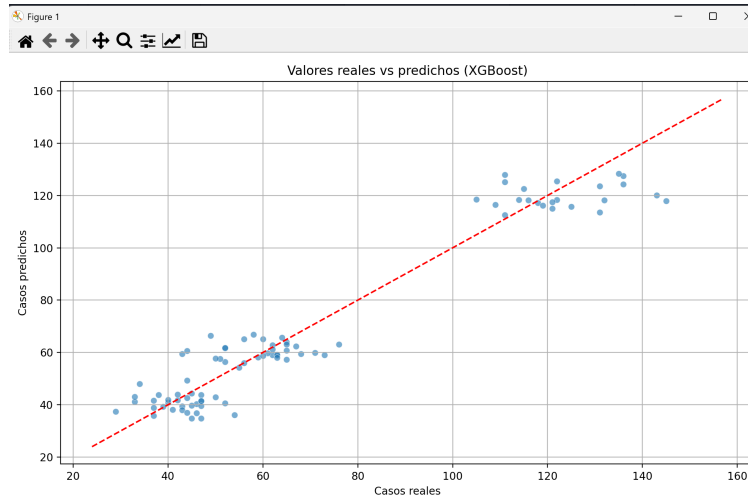


Figure 1: Predicción con XGBoost (sin optimización)

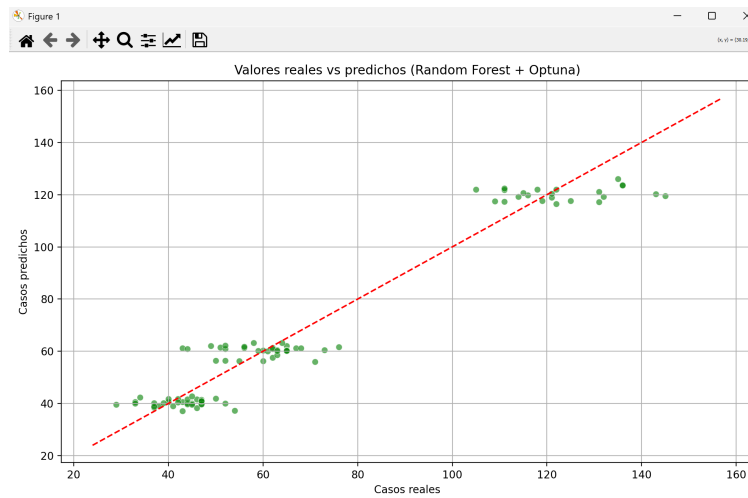


Figure 2: Predicción con Random Forest optimizado con Optuna

Modelo	MSE	R^2
XGBoost	77.84	0.93
Random Forest + Optuna	[MSE_final]	[R2_final]

Table 1: Comparación de resultados entre modelos

5 Conclusiones

Los modelos de regresión permiten una buena estimación de los casos diarios de COVID-19. La optimización con Optuna mejoró significativamente el desempeño del modelo Random Forest, mostrando mayor alineación con los valores reales.

6 Diagrama de Flujo

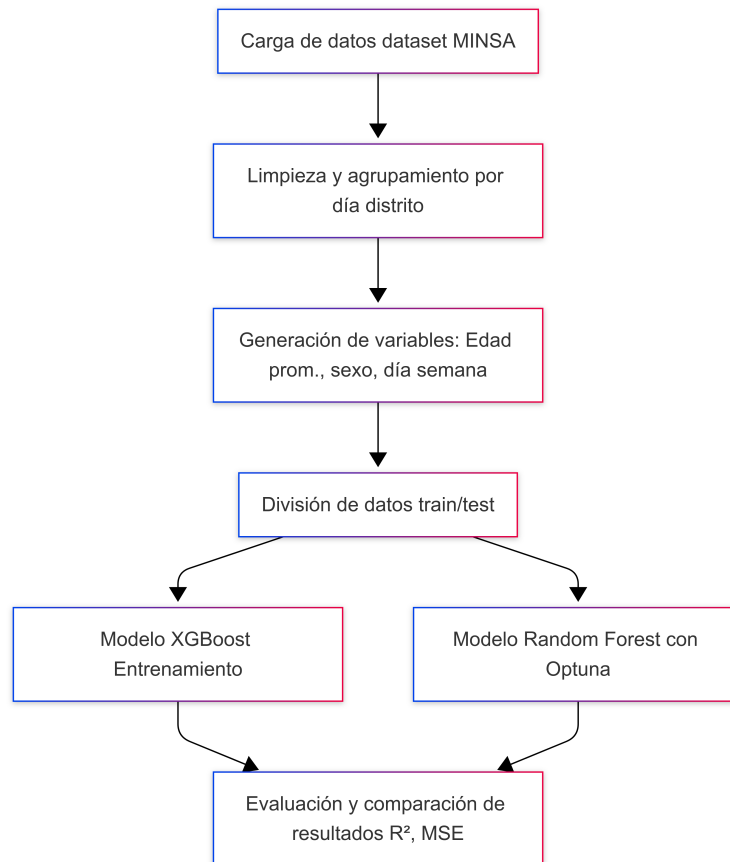


Figure 3: Flujo del proceso desde el preprocesamiento hasta la evaluación

7 Bibliografía

- Ministerio de Salud del Perú. Casos positivos por COVID-19. Plataforma de Datos Abiertos.
<https://www.datosabiertos.gob.pe/>
- Scikit-learn: <https://scikit-learn.org/>
- Optuna: <https://optuna.org/>
- XGBoost: <https://xgboost.readthedocs.io/>