



UNIVERSIDAD
COMPLUTENSE
MADRID

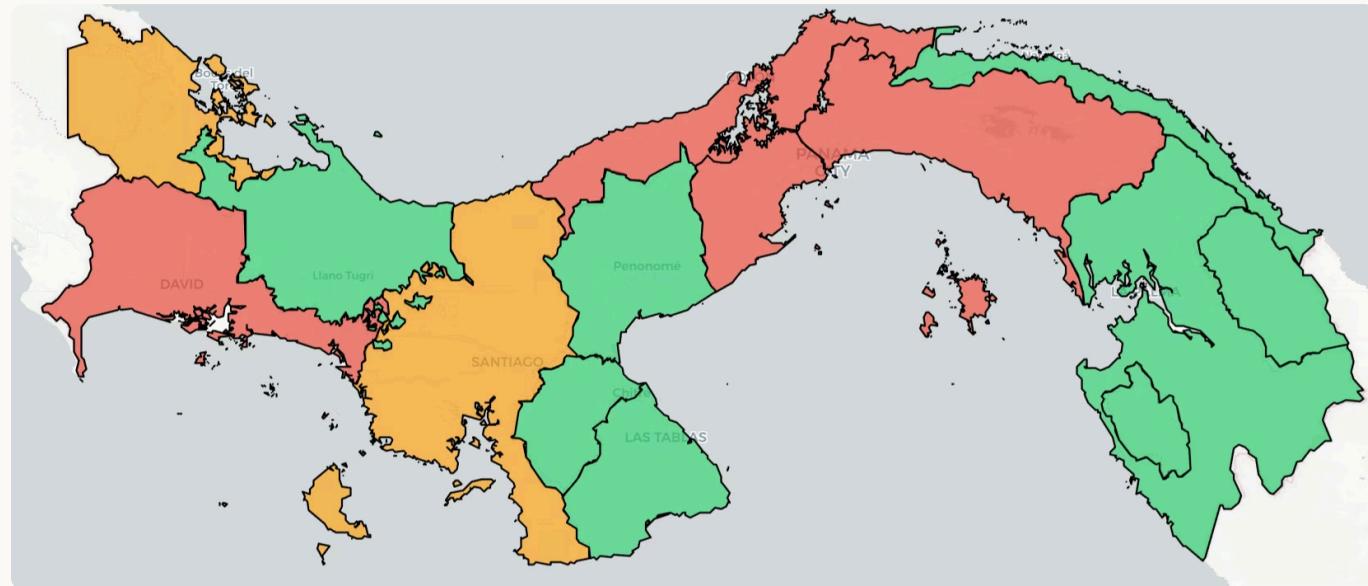
TFM - APPLICACIÓN DE MACHINE LEARNING PARA LA
PREDICCIÓN DE DESEMPLÉO REGIONAL EN PANAMÁ: UN
ENFOQUE DE POLÍTICA PÚBLICA BASADO EN DATOS Y
ANÁLISIS GEOESPACIAL

Henry De Gracia

Máster Data Science, Big Data & Business Analytics
Universidad Complutense de Madrid, 2026



El Problema: Brechas Estructurales en el Mercado Laboral



A quien va dirigido:

- Ministerio de Trabajo de Panama
- Planificadores de politica publica regionales
- Investigadores del mercado laboral centroamericano

- **Objetivo del Proyecto:** Construir un modelo predictivo de tasa de desempleo por provincia y clasificar niveles de riesgo laboral para apoyar políticas públicas.

El proyecto transforma datos publicos dispersos en una herramienta predictiva e interpretable que permite anticipar crisis de desempleo por provincia, entender sus causas, y simular el impacto de intervenciones antes de implementarlas.

Desigualdades Significativas

Panama carece de herramientas predictivas para anticipar el desempleo a nivel regional. Las estadisticas oficiales del INEC se publican con meses de retraso y solo describen lo que ya paso, sin capacidad de proyectar ni alertar sobre crisis futuras.

Impacto de la Pandemia

El desempleo nacional se dispara al 18.5% en 2020

La recuperacion fue desigual: algunas provincias se recuperaron rapido, otras aun no lo logran en 2024

Esto evidencio que no habia herramientas para anticipar que provincias serian mas vulnerables

Fuentes de Datos



Fuentes Oficiales

[59 archivos Excel del INEC Panamá con encuestas de mercado laboral.](#)

Variables de Contexto Macro

[8 indicadores macroeconómicos del Banco Mundial.](#)

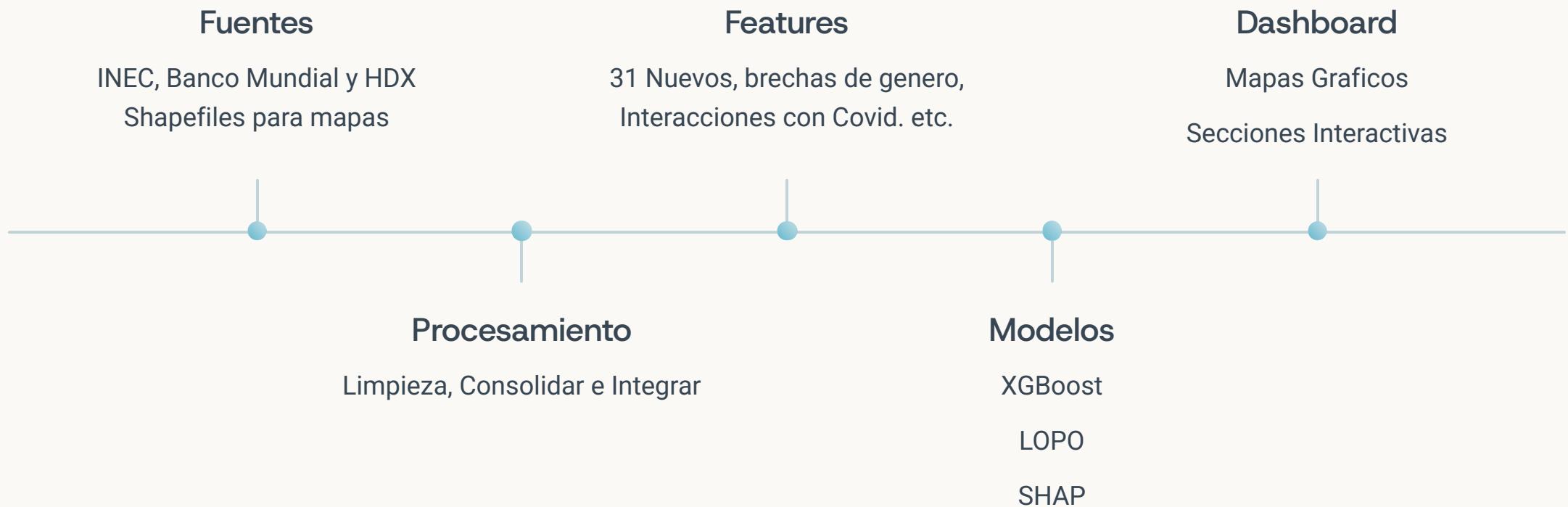
Geoespaciales

[Shapefiles geoespaciales de provincias de Panamá](#)

Dataset Consolidado

- 1,605 observaciones
- 13 provincias y comarcas
- 17 periodos (2011-2024)
- Desagregación por área y sexo
- 31 features creadas
- Variables: retardos temporales, brechas de género, interacciones con COVID

Pipeline Analítico Del Proyecto



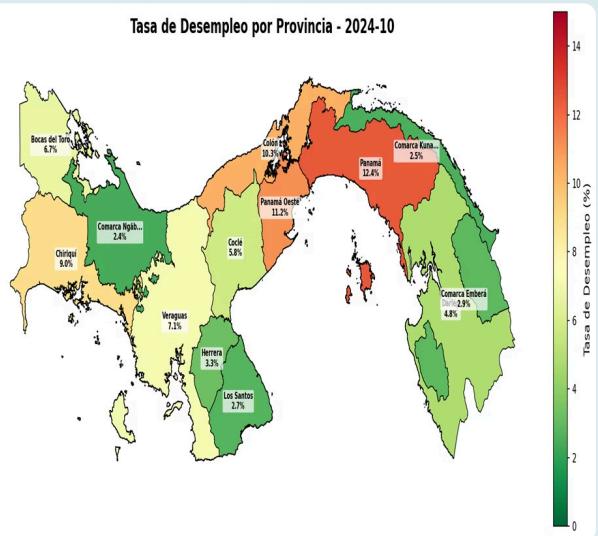
1. Fuentes → Se recolectaron datos de 3 fuentes: 59 archivos Excel del INEC con encuestas laborales, indicadores macroeconomicos del Banco Mundial via API, y shapefiles de HDX para los mapas.
2. Procesamiento → Se limpiaron y consolidaron los datos en un unico dataset de 1,605 observaciones (13 provincias x 17 periodos x area x sexo), estandarizando nombres y corrigiendo inconsistencias.
3. Features → Se crearon 31 variables predictivas: retardos temporales (desempleo del periodo anterior), cambios entre periodos, brechas de genero, interacciones con COVID, y variables categoricas de provincia/area/sexo.
4. Modelo → Se entrenaron y compararon 5 algoritmos con validacion cruzada temporal (LOPO). XGBoost fue el mejor con $R^2=0.77$ y precision de clasificacion de riesgo del 79.4%. Se añadio interpretabilidad con SHAP.
5. Dashboard → Se productivizo el modelo en una aplicacion Streamlit con 8 secciones: mapas interactivos, graficos temporales, analisis de riesgo, explicaciones SHAP, y un predictor que permite hacer simulaciones en tiempo real.

Análisis Exploratorio

Evolucion del Desempleo por Provincia (2011-2024)



Heterogeneidad Provincial



Existe una brecha significativa entre provincias. Colon y las comarcas indigenas presentan tasas consistentemente superiores al 10%, mientras que Los Santos y Herrera se mantienen por debajo del 5%.

Recuperación Desigual

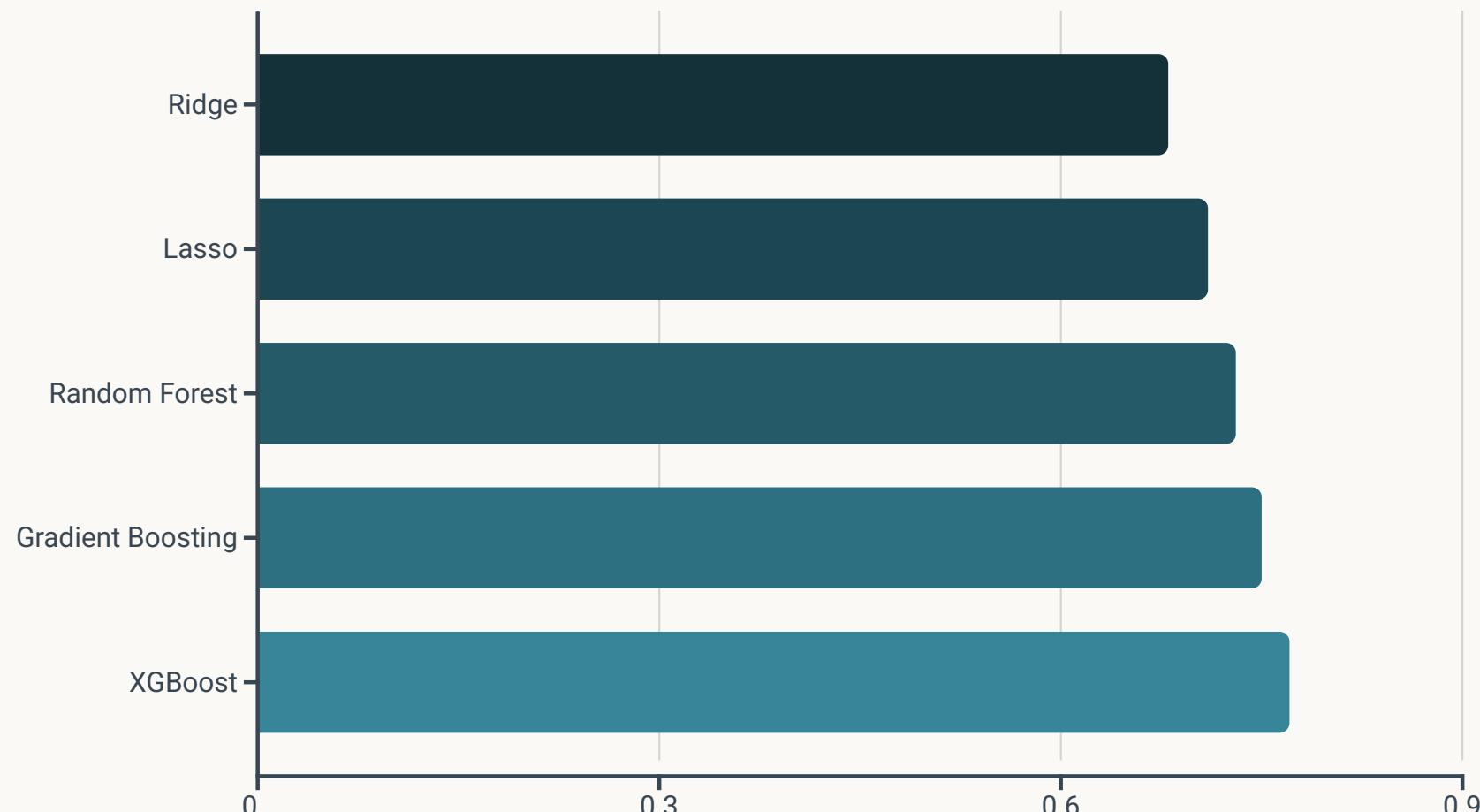
- Impacto COVID-19:** el desempleo promedio aumento +2.3 pp tras la pandemia (de 3.7% a 5.9%). Panamá Oeste fue la provincia mas afectada (+7.6 pp). La recuperacion fue desigual entre provincias.
- Se observa un pico generalizado en 2020 (encuesta telefonica), seguido de una recuperacion gradual pero desigual. Las provincias con mayor informalidad laboral sufrieron caidas mas pronunciadas.

Brechas

- Brecha de genero:** las mujeres presentan un desempleo promedio de 5.4% vs 3.4% de los hombres (+2.0 pp), persistente en todos los periodos.
- Brecha urbano-rural:** el desempleo urbano (6.1%) supera al rural (4.2%) en +1.9 pp, porque en areas rurales el desempleo se oculta en informalidad y trabajo de subsistencia.

Modelado y Evaluación

Comparación de Algoritmos



0.77

R-cuadrado

Explica 77% de variabilidad

1.13

Error Medio

Puntos porcentuales

79.4%

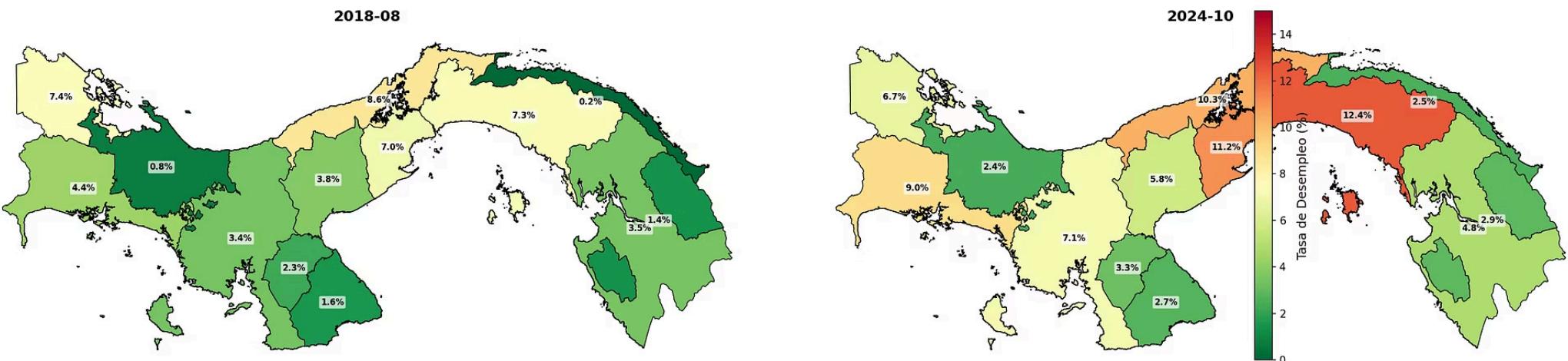
Precisión de Clasificación

4 de cada 5 casos correctos

- Validación: Leave-One-Period-Out - simula escenario real de predicción temporal

Análisis Geoespacial: Patrones y Discrepancias

Evolucion del Desempleo por Provincia

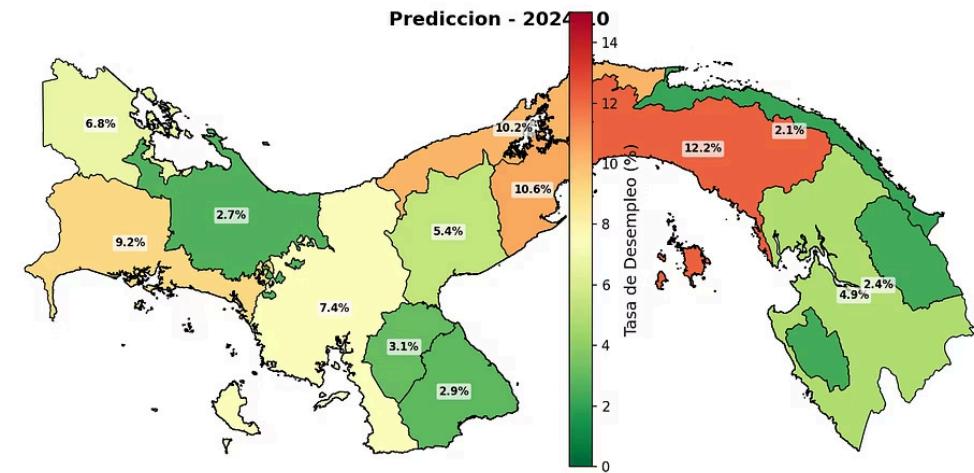
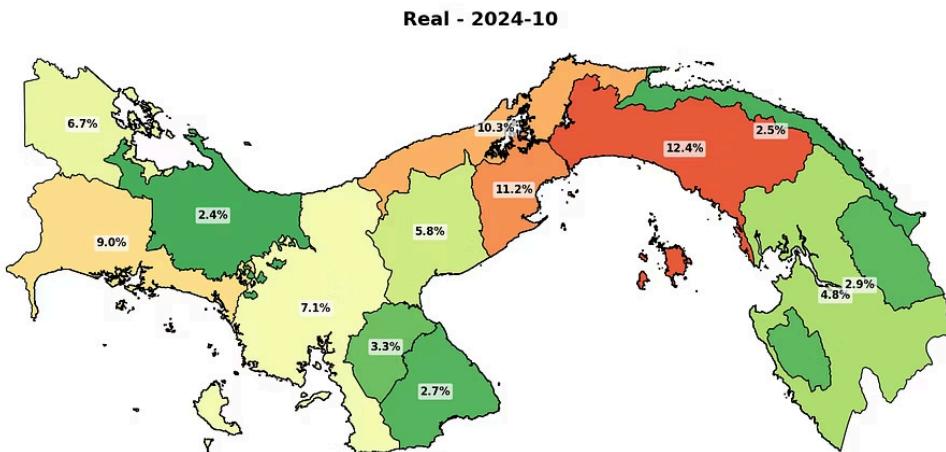


Evolución Temporal

Comparación 2018 vs 2024: permite evaluar mejora o deterioro provincial

Análisis Geoespacial: Patrones y Discrepancias

Desempleo Real vs Prediccion del Modelo



Real vs Predicción

Captura patrón geográfico correctamente. Discrepancias menores en comarcas con datos limitados

Dashboard Interactivo: Productivización

1 Resumen del Modelo

2 Mapa Interactivo

3 Evolucion Temporal

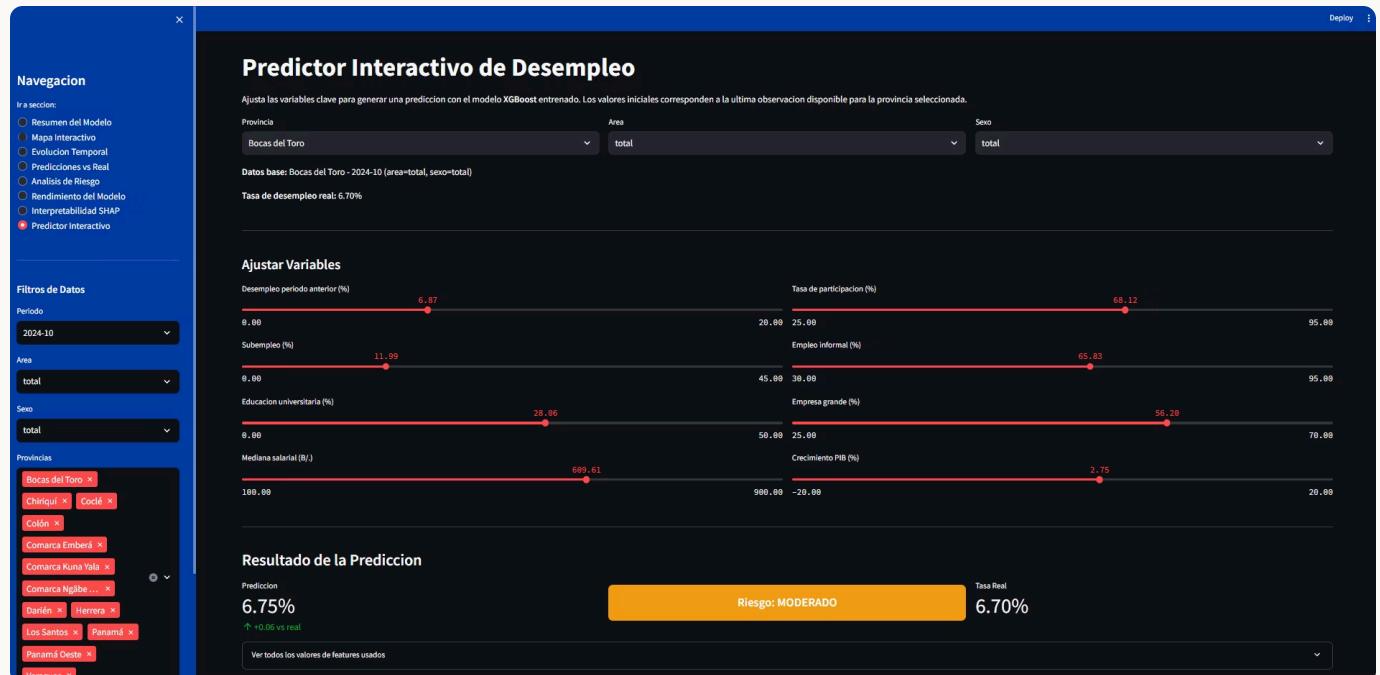
4 Predicciones vs Real

5 Analisis de Riesgo

6 Rendimiento del
Modelo

7 Interpretabilidad SHAP

8 Predictor Interactivo



8 Secciones + 80 Tests

El predictor interactivo permite análisis tipo "qué pasaría si": ajustar variables y generar predicción con clasificación de riesgo

- Ejemplo: Si el subempleo en Colón sube del 20% al 30%, el modelo predice aumento significativo en desempleo, pasando a riesgo crítico

Conclusiones y Lecciones Aprendidas

Principales Conclusiones

Predicción

Es posible predecir desempleo regional con datos públicos, explicando 77% de variabilidad

Subempleo como Predictor

Mejor predictor del desempleo futuro
- políticas deberían enfocarse en calidad del empleo

Inercia del Desempleo

Fuerte inercia justifica intervenciones tempranas

Lecciones Clave



Calidad de Datos

Mayor reto: formatos inconsistentes entre archivos Excel del INEC



Validación Temporal

Esencial para no sobreestimar rendimiento



Valor de Productivización

Dashboard diferenciador: ejercicio académico vs. herramienta útil

Trabajo Futuro y Cierre



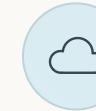
Extensión Geográfica

Modelo a nivel de distrito o corregimiento en lugar de provincial



Automatización Automática

Pipeline que descargue nuevos datos del INEC automáticamente y re-entrene el modelo



Despliegue en Nube

Dashboard accesible online para tomadores de decisión

Muchas gracias

Henry De Gracia | Master en Data Science, de Big Data & Business Analytics | Universidad Complutense de Madrid, 2026

github.com/henryddr/TFM-Desempleo-Panama