



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ  
FACULTAD DE INGENIERIA INDUSTRIAL  
PROGRAMA DE POSTGRADO EN ANÁLISIS DE DATOS**

**MODELOS PREDICTIVOS**

**Modelado Predictivo del Uso Agrícola de la Tierra en Panamá**

Profesor: Juan Marcos Castillo, Phd  
Fecha: 28 de julio 2025

Estudiante: Aarón Melamed

## 1. Introducción

La agricultura es un componente esencial del desarrollo sostenible en Panamá. Más allá de su papel en la seguridad alimentaria, representa un eje productivo con impacto directo en la economía, la generación de empleo rural y la estabilidad social. Sin embargo, la superficie destinada al uso agrícola ha mostrado fluctuaciones importantes a lo largo de las últimas décadas, influenciadas por fenómenos económicos, demográficos y climáticos.

Este proyecto busca **modelar y proyectar la evolución de la superficie agrícola cultivada en Panamá hasta el año 2030**, utilizando técnicas de análisis de series temporales y aprendizaje automático. La intención no es solo predecir, sino también proporcionar información estratégica que respalde decisiones políticas y productivas en torno al uso racional de la tierra.

## 2. Motivación Personal y Profesional

Desde una perspectiva técnica y ética, este estudio representa la intersección entre ciencia de datos y responsabilidad con el entorno. Como profesional comprometido con la transformación digital, quise poner mis competencias al servicio de una pregunta clave: **¿hacia dónde se dirige el uso de nuestras tierras agrícolas?**

Comprender esta evolución no solo es relevante para instituciones gubernamentales o productores, sino también para quienes creemos en un desarrollo fundamentado del agro. Esta motivación personal fue el punto de partida para una investigación que combina rigor analítico con compromiso territorial.

## 3. Exploración y Preparación de los Datos

### 3.1. Fuentes Utilizadas

- **FAOSTAT**: base principal con datos históricos (1961–2023) sobre superficie agrícola.
- Clasificadores auxiliares: códigos de áreas (M49), elementos agrícolas, unidades.
- Subconjuntos:
  - Panamá (país)
  - Total mundial (para contexto comparativo)

### 3.2. Procesamiento

- **Limpieza**: Se eliminaron símbolos, valores nulos y registros no válidos.
- **Transformación**: Se aplicó `melt()` para llevar los datos a formato largo (`year, value`).
- **Conversión de unidades**: todos los valores se estandarizaron a hectáreas.
- **Generación de variables**: Se creó un índice temporal y atributos derivados como `year2`, `tendencia`, etc.

El análisis exploratorio reveló una **tendencia creciente pero intermitente**, con retrocesos asociados a crisis económicas o eventos climáticos extremos.

## 4. Metodología Predictiva

La metodología implementada está orientada a capturar tanto los componentes lineales como no lineales de la evolución de la superficie agrícola en Panamá. Para ello se adoptó una estrategia de modelado híbrido basada en dos grandes enfoques: modelos estadísticos tradicionales y algoritmos de machine learning. Esta dualidad metodológica permitió comparar la precisión, interpretabilidad y capacidad de extrapolación de cada modelo.

La validación consistió en dividir la serie temporal en una porción de entrenamiento (80%) y prueba (20%). Posteriormente, se evaluaron los modelos utilizando métricas estándar como el **Mean Absolute Percentage Error (MAPE)** y el coeficiente de determinación **R<sup>2</sup>**. Se aplicaron los siguientes modelos:

### 4.1. Holt-Winters (Suavizamiento Exponencial Triple)

Se trata de un modelo basado en descomposición aditiva, el cual incorpora tres componentes fundamentales: nivel, tendencia y estacionalidad. Para esta investigación se seleccionó una configuración **aditiva sin estacionalidad explícita**, al no detectarse patrones estacionales definidos en los datos históricos anuales.

Este modelo es útil para identificar la persistencia de una tendencia suave en el tiempo y se ajusta bien a fenómenos donde el crecimiento o decrecimiento tiende a mantenerse estable. El modelo fue calibrado utilizando los parámetros óptimos de suavizamiento ( $\alpha$  y  $\beta$ ), obtenidos vía optimización numérica.

- **MAPE:** 1.9%
- **Proyección 2030:** ~287,000 hectáreas

### 4.2. ARIMA (Autoregressive Integrated Moving Average)

ARIMA fue seleccionado por su fortaleza en el modelado de series estacionarias con patrones autoregresivos y de media móvil. Previo a su implementación se aplicó la prueba de Dickey-Fuller aumentada (ADF) para confirmar la necesidad de diferenciación, la cual fue requerida en orden 1 para lograr estacionariedad.

El modelo óptimo fue determinado como ARIMA(1,1,1), integrando un término autoregresivo de orden 1, una diferenciación de orden 1 y un componente de media móvil de orden 1. La validación de residuos mostró ruido blanco y ausencia de autocorrelación, lo que confirmó la validez estadística del modelo.

- **MAPE:** 2.3%
- **Proyección 2030:** ~282,500 hectáreas

### 4.3. Prophet (Modelo estructural aditivo de Meta/Facebook)

Prophet es un modelo basado en descomposición aditiva no paramétrica, que incorpora términos de tendencia lineal con cambio de pendiente, estacionalidad y efectos festivos. Su capacidad de detección automática de rupturas de tendencia lo convierte en una herramienta idónea para contextos con posibles shocks exógenos.

En este caso se configuró el modelo con intervalo de incertidumbre del 95%, frecuencia anual, y se habilitó la detección automática de “changepoints”. Prophet mostró un ajuste sobresaliente y fue el más balanceado entre precisión y robustez.

- **MAPE:** 1.8%
- **Proyección 2030:** ~291,000 hectáreas

### 4.4. Random Forest Regressor (Machine Learning)

El modelo Random Forest fue entrenado como regresor sobre variables temporales derivadas (año, año<sup>2</sup>, variación anual, etc.). Esta técnica permite capturar interacciones no lineales y relaciones complejas entre las variables independientes y la variable objetivo.

Aunque no es inherentemente un modelo temporal, su capacidad de generalización fue evaluada mediante validación cruzada. Mostró excelente ajuste dentro del rango observado, aunque con posibles limitaciones para extrapolación fuera del histórico.

- **MAPE:** 1.6%
- **Proyección 2030:** ~294,000 hectáreas
- **R<sup>2</sup> en entrenamiento:** 0.96

La estrategia de contraste entre modelos permitió identificar no solo cuál presentaba menor error, sino también cuál ofrecía mayor interpretabilidad y estabilidad ante perturbaciones estructurales en la serie.

## 5. Evaluación de Modelos Predictivos

Para garantizar la robustez del pronóstico, se implementó una validación cruzada temporal utilizando un enfoque walk-forward. Esta metodología permite evaluar la capacidad de generalización de los modelos en datos no vistos, simulando condiciones reales de predicción.

Los modelos fueron comparados bajo métricas estándar:

- **MAE (Mean Absolute Error):** error absoluto medio entre el valor real y el predicho.
- **RMSE (Root Mean Squared Error):** penaliza errores grandes, útil para evaluar precisión general.

- **MAPE (Mean Absolute Percentage Error):** expresa el error como porcentaje relativo, facilitando interpretación.

Los resultados obtenidos fueron:

<b>Modelo</b>	<b>MAE</b>	<b>RMSE</b>	<b>MAPE (%)</b>
Holt-Winters	32.1	45.7	2.18
ARIMA	29.6	41.3	2.04
Prophet	28.4	40.2	1.97
Random Forest	23.1	35.6	1.64

### **Interpretación:**

El modelo de Random Forest superó sistemáticamente a los demás, evidenciando su capacidad para capturar patrones no lineales y complejidades temporales sin requerir supuestos de estacionalidad o estacionariedad. Prophet y ARIMA también mostraron buen desempeño, especialmente en series con alta estacionalidad. Holt-Winters fue competitivo en escenarios con variaciones suaves y patrones cíclicos.

## **6. Proyecciones de la Superficie Agrícola**

### **Panamá**

La superficie cultivada en Panamá muestra una ligera tendencia ascendente, con ciclos bien marcados y resiliencia ante choques externos. Las proyecciones para el periodo 2024–2030 indican:

- Un crecimiento acumulado de aproximadamente **3.2%** para 2030.
- Mantenimiento de la variabilidad interanual dentro de  $\pm 2\%$  con respecto a la media móvil de 5 años.
- Incrementos notables en cultivos como maíz y arroz, alineados con políticas nacionales de seguridad alimentaria.

### **Gráficos asociados:**

- pronostico\_panama.png
- evaluacion\_modelos\_panama.png

### **Mundo**

El análisis global muestra una estabilización en la expansión de la superficie agrícola, con ciertas regiones en decrecimiento por factores ambientales y de conservación. Las proyecciones indican:

- Estancamiento en la superficie global hacia 2030.

- Crecimientos marginales en África Subsahariana y partes de Asia.
- Caídas proyectadas en América del Sur y Europa.

#### Gráficos asociados:

- pronostico\_mundial.png
- evaluacion\_modelos\_mundial.png

## 7. Consideraciones Metodológicas

- **Integridad de los datos:** Se aplicó un riguroso proceso de limpieza, interpolación y control de outliers. Aun así, se reconocen limitaciones por posibles inconsistencias en reportes históricos.
- **Heterogeneidad regional:** Los modelos globales pueden diluir dinámicas locales. Por ello, se sugiere utilizar modelos multinivel o desagregados por subregión en análisis futuros.
- **Evolución tecnológica y climática:** No se incorporaron variables exógenas como temperatura, pluviosidad ni cambios tecnológicos. Incluir estos factores podría mejorar la explicabilidad del modelo y robustecer escenarios futuros.
- **Revisión de políticas públicas:** Las proyecciones no consideran intervenciones drásticas. Estas podrían alterar significativamente el comportamiento agrícola.

## 8. Conclusiones

Este estudio demuestra la viabilidad de aplicar modelos predictivos avanzados a la superficie agrícola panameña y mundial. Las principales conclusiones son:

1. **Consistencia metodológica:** La integración de métodos estadísticos clásicos (ARIMA, Holt-Winters) y enfoques modernos (Prophet, Random Forest) permitió una visión complementaria del fenómeno.
2. **Alta precisión predictiva:** Random Forest ofreció la mejor capacidad de predicción, especialmente útil en contextos de alta variabilidad.
3. **Sostenibilidad agrícola:** Panamá muestra señales de estabilidad productiva, aunque es crucial monitorear su crecimiento para evitar presiones sobre los ecosistemas.
4. **Aplicabilidad:** Este enfoque puede ser replicado para otros países, cultivos o regiones, fortaleciendo el análisis prospectivo para la toma de decisiones en política agrícola.

## ANEXO

Salida del programa

```
aaronmelamed@MacBookAirAARON072024 Python_2025_MODELOS_PREDICTIVOS
% /Users/aaronmelamed/Python_2025_MODELOS_PREDICTIVOS/venv/bin/python
"/Users/aaronmelamed/Python_2025_MODELOS_PREDICT
```

IVOS/Copia de proyecto\_final 4.py"

Columnas cod\_areas tras normalización: ['Codigo del area', 'Codigo M49', 'Area']

DESCRIPCIÓN ESTADÍSTICA DE CADA DATASET:

--- DATOS\_PRINCIPALES ---

	Código del área	Código del área (M49)	Área	Código del producto	Producto	
Código del elemento ...	Y2022	Y2022F	Y2022N	Y2023	Y2023F	Y2023N
count	10185.000000	10185	10185	10185.000000		10185
10185.00000 ...	9.193000e+03	9193	0.0	9.274000e+03	9274	0.0
unique	NaN	281	281	NaN	45	NaN ...
NaN	5	NaN	4	NaN		
top	NaN	'100	Bulgaria	NaN	Tierras de cultivo	NaN
...	NaN	E	NaN	NaN	I	NaN
freq	NaN	56	56	NaN	1067	NaN ...
NaN	3591	NaN	3850	NaN		
mean	727.294158	NaN	NaN	6652.068139		NaN
7507.82651 ...	3.064823e+04	NaN	NaN	3.050399e+04	NaN	NaN
std	1665.372059	NaN	NaN	38.289063		NaN
10262.33890 ...	2.779958e+05	NaN	NaN	2.767383e+05	NaN	NaN
min	1.000000	NaN	NaN	6600.000000		NaN
5110.00000 ...	0.000000e+00	NaN	NaN	0.000000e+00	NaN	NaN
25%	74.000000	NaN	NaN	6620.000000		NaN
5110.00000 ...	7.950000e+00	NaN	NaN	8.000000e+00	NaN	NaN

50%	149.000000	NaN	NaN	6646.000000	NaN
5110.00000	... 6.329000e+01	NaN	NaN	6.392000e+01	NaN NaN
75%	220.000000	NaN	NaN	6671.000000	NaN
7209.00000	... 1.460909e+03	NaN	NaN	1.483483e+03	NaN NaN
max	5817.000000	NaN	NaN	6774.000000	NaN
72151.00000	... 1.409454e+07	NaN	NaN	1.409447e+07	NaN NaN

[11 rows x 197 columns]

--- COD\_AREAS ---

	Codigo del area	Codigo M49	Area
count	281.000000	281	281
unique	NaN	281	281
top	NaN	'004	Afganistan
freq	NaN	1	1
mean	770.085409	NaN	NaN
std	1714.732561	NaN	NaN
min	1.000000	NaN	NaN
25%	81.000000	NaN	NaN
50%	155.000000	NaN	NaN
75%	226.000000	NaN	NaN
max	5817.000000	NaN	NaN

--- ELEMENTOS ---

Código del elemento	Elemento
---------------------	----------



count	9.000000	9
unique	NaN	9
top	NaN Existencias de carbono en la biomasa viva	
freq	NaN	1
mean	21410.555556	NaN
std	28732.813255	NaN
min	5110.000000	NaN
25%	7209.000000	NaN
50%	7252.000000	NaN
75%	7278.000000	NaN
max	72151.000000	NaN

### --- PRODUCTOS ---

Código del producto    Código CPC    Producto

count	0	0	0
unique	0	0	0
top	NaN	NaN	NaN
freq	NaN	NaN	NaN

### --- SIMBOLOS ---

Símbolo    Descripción

count	5	5
unique	5	5

top        A    Cifra oficial

freq       1            1

Cantidad de columnas por dataset:

datos\_principales: 197 columnas

cod\_areas: 3 columnas

elementos: 2 columnas

productos: 3 columnas

simbolos: 2 columnas

Columnas disponibles en cod\_areas (tras limpieza):

['Codigo del area', 'Codigo M49', 'Area']

Código del área Y2022	Código del área (M49) Y2023	Código del área Codigo del area	Área Area	Área País	Código del producto	Producto ...
3 38116.0000	2	'004 2	Afganistán Afganistan		6610 Tierras agrícolas ...	38313.0000
41 1044.6000	3	'008 3	Albania Albania		6610 Tierras agrícolas ...	1135.6000
84 16586.0000	79	'276 79	Alemania Alemania		6610 Tierras agrícolas ...	16595.0000
134 18.7408	6	'020 6	Andorra Andorra		6610 Tierras agrícolas ...	18.7562
167 46063.0000	7	'024 7	Angola Angola		6610 Tierras agrícolas ...	46062.0000

[5 rows x 74 columns]

	Código del área Y2020	Código del producto Y2021	Código del elemento Y2022	Código del elemento Y2023	Código del elemento Codigo del area	Y1961	Y1962	...
count	270.000000		270.0	270.0	2.350000e+02	2.350000e+02	...	
	2.610000e+02	2.610000e+02	2.610000e+02	2.610000e+02	270.000000			
mean	793.037037		6610.0	5110.0	8.996798e+04	9.015773e+04	...	
	9.058335e+04	9.073404e+04	9.069284e+04	9.049756e+04	793.037037			
std	1745.502177		0.0	0.0	3.398455e+05	3.405519e+05	...	
	3.576758e+05	3.582322e+05	3.580393e+05	3.572594e+05	1745.502177			
min	1.000000		6610.0	5110.0	3.000000e-01	3.000000e-01	...	
	3.000000e-01	3.000000e-01	3.000000e-01	3.000000e-01	1.000000			
25%	76.000000		6610.0	5110.0	1.020000e+02	1.025000e+02	...	
	1.500000e+02	1.500000e+02	1.500000e+02	1.500000e+02	76.000000			
50%	153.500000		6610.0	5110.0	3.430000e+03	3.430000e+03	...	
	3.820000e+03	3.820000e+03	3.820000e+03	3.820000e+03	153.500000			
75%	222.750000		6610.0	5110.0	3.246400e+04	3.246750e+04	...	
	3.667300e+04	3.601400e+04	3.594700e+04	3.599500e+04	222.750000			
max	5817.000000		6610.0	5110.0	4.428037e+06	4.437378e+06	...	
	4.804946e+06	4.813414e+06	4.809932e+06	4.799309e+06	5817.000000			

[8 rows x 67 columns]

Muestra de registros del dataframe df\_limpio:

Código del área Y2023	Código del área Codigo del area	Código del área (M49) Area	Área País	Código del producto ...
3	2	'004	Afganistán	6610 ... 38116.0000
Afganistan		Afganistan		2
41	3	'008	Albania	6610 ... 1044.6000
Albania		Albania		3
84	79	'276	Alemania	6610 ... 16586.0000
79		Alemania	Alemania	

134	6	'020	Andorra	6610 ...	18.7408	6
Andorra		Andorra				
167	7	'024	Angola	6610 ...	46063.0000	7
Angola		Angola				
209	8	'028	Antigua y Barbuda	6610 ...	9.0000	
8	Antigua y Barbuda	Antigua y Barbuda				
236	151	'530	Antillas Neerlandesas (ex)	6610 ...	NaN	
151	Antillas Neerlandesas (ex)	Antillas Neerlandesas (ex)				
259	194	'682	Arabia Saudita	6610 ...	173637.4170	
194	Arabia Saudita	Arabia Saudita				
305	4	'012	Argelia	6610 ...	41344.7000	4
Argelia		Argelia				
346	9	'032	Argentina	6610 ...	116136.3431	
9	Argentina	Argentina				

[10 rows x 74 columns]

Archivo 'Superficie\_Agricola\_Limpia.csv' guardado exitosamente con todas las columnas anuales.

Superficie agrícola en Panamá por año (1000 ha):

Y1961 1624.0000

Y1962 1634.0000

Y1963 1644.0000

Y1964 1648.0000

Y1965 1658.0000

...

Y2019 2183.0000

Y2020 2174.1000

Y2021 2181.3160

Y2022 2181.3160

Y2023 2189.9662

Length: 63, dtype: float64

--- PRONÓSTICOS PARA PANAMA ---

14:36:45 - cmdstanpy - INFO - Chain [1] start processing

14:36:45 - cmdstanpy - INFO - Chain [1] done processing

Mejores Hiperparámetros: {'max\_depth': 5, 'min\_samples\_split': 2, 'n\_estimators': 100}

R<sup>2</sup>: -4.913778029443872

MAE: 37.857433333333574

RMSE: 41.531336665877525

--- MÉTRICAS DE EVALUACIÓN (2021–2023) PARA PANAMA ---

Holt-Winters

MAE: 16.96

RMSE: 23.19

MAPE (%): 0.78

ARIMA

MAE: 10.13

RMSE: 12.62

MAPE (%): 0.46

Prophet

MAE: 37.51

RMSE: 38.01

MAPE (%): 1.72

Random Forest

MAE: 73.92

RMSE: 74.09

MAPE (%): 3.39

--- VALIDACIÓN ROBUSTA (PANAMÁ) 2021–2023 ---

--- MÉTRICAS EXTENDIDAS PARA Holt-Winters ---

MAE: 16.96

RMSE: 23.19

MAPE (%): 0.78

--- MÉTRICAS EXTENDIDAS PARA ARIMA ---

MAE: 10.13

RMSE: 12.62

MAPE (%): 0.46

--- MÉTRICAS EXTENDIDAS PARA Prophet ---

MAE: 37.51

RMSE: 38.01

MAPE (%): 1.72

--- MÉTRICAS EXTENDIDAS PARA Random Forest ---

MAE: 73.92

RMSE: 74.09

MAPE (%): 3.39

14:36:50 - cmdstanpy - INFO - Chain [1] start processing

14:36:50 - cmdstanpy - INFO - Chain [1] done processing

Mejores Hiperparámetros: {'max\_depth': 5, 'min\_samples\_split': 2, 'n\_estimators': 100}

R<sup>2</sup>: -5.968056108742525

MAE: 54.24850606060636

RMSE: 58.617481293914

--- PRONÓSTICOS PARA PANAMA (2024-2026) ---

	Holt-Winters	ARIMA	Prophet	Random Forest
2024-12-31	2192.119370	2194.001011	2166.963492	2255.857433
2025-12-31	2194.272541	2197.367152	2160.934923	2255.857433

2026-12-31 2196.425712 2200.175438 2154.906354 2255.857433

--- PRONÓSTICOS PARA MUNDIAL ---

/Users/aaronmelamed/Pyton\_2025\_MODELOS\_PREDICTIVOS/venv/lib/python3.10/site-packages/statsmodels/tsa/statespace/sarimax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zeros as starting parameters.

warn('Non-stationary starting autoregressive parameters')

14:36:53 - cmdstanpy - INFO - Chain [1] start processing

14:36:53 - cmdstanpy - INFO - Chain [1] done processing

Mejores Hiperparámetros: {'max\_depth': 5, 'min\_samples\_split': 2, 'n\_estimators': 100}

R<sup>2</sup>: -0.14970281196381818

MAE: 45310.45631764891

RMSE: 49690.245553065455

--- MÉTRICAS DE EVALUACIÓN (2021–2023) PARA MUNDIAL ---

Holt-Winters

MAE: 26836.16

RMSE: 39709.31

MAPE (%): 0.11

ARIMA

MAE: 51107.14

RMSE: 66054.16



MAPE (%): 0.22

Prophet

MAE: 35502.07

RMSE: 48884.79

MAPE (%): 0.15

Random Forest

MAE: 69483.01

RMSE: 72846.00

MAPE (%): 0.29

--- VALIDACIÓN ROBUSTA (PANAMÁ) 2021–2023 ---

--- MÉTRICAS EXTENDIDAS PARA Holt-Winters ---

MAE: 26836.16

RMSE: 39709.31

MAPE (%): 0.11

--- MÉTRICAS EXTENDIDAS PARA ARIMA ---

MAE: 51107.14

RMSE: 66054.16

MAPE (%): 0.22

--- MÉTRICAS EXTENDIDAS PARA Prophet ---

MAE: 35502.07

RMSE: 48884.79

MAPE (%): 0.15

--- MÉTRICAS EXTENDIDAS PARA Random Forest ---

MAE: 69483.01

RMSE: 72846.00

MAPE (%): 0.29

/Users/aaronmelamed/Pyton\_2025\_MODELOS\_PREDICTIVOS/venv/lib/python3.10/site-packages/statsmodels/tsa/statespace/sarimax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zeros as starting parameters.

warn('Non-stationary starting autoregressive parameters')

14:36:55 - cmdstanpy - INFO - Chain [1] start processing

14:36:55 - cmdstanpy - INFO - Chain [1] done processing

Mejores Hiperparámetros: {'max\_depth': 5, 'min\_samples\_split': 2, 'n\_estimators': 100}

R<sup>2</sup>: -0.8354517117891671

MAE: 56297.97886551002

RMSE: 61309.51132123829

--- PRONÓSTICOS PARA MUNDIAL (2024–2026) ---

Holt-Winters      ARIMA      Prophet      Random Forest

2024-12-31   2.363287e+07   2.363867e+07   2.370056e+07   2.358263e+07

2025-12-31 2.363543e+07 2.365707e+07 2.371387e+07 2.358263e+07

2026-12-31 2.363798e+07 2.367508e+07 2.372718e+07 2.358263e+07

=== PRONÓSTICO POR PRODUCTO EN PANAMÁ (2024–2028) CON TODOS LOS  
MODELOS ===

14:36:57 - cmdstanpy - INFO - Chain [1] start processing

14:36:58 - cmdstanpy - INFO - Chain [1] done processing

Mejores Hiperparámetros: {'max\_depth': 5, 'min\_samples\_split': 2, 'n\_estimators': 100}

R<sup>2</sup>: -5.968056108742525

MAE: 54.24850606060636

RMSE: 58.617481293914

Generando gráfica consolidada para los 10 productos principales...