

Proyecto Final

Bootcamp IA 2025 - Unicuses

Análisis y Modelado de Cultivos con Machine Learning

Presentado por:

Diego Fernando Vélez Neira



Objetivo del Proyecto

- Aplicar técnicas de Machine Learning para analizar y predecir la superficie sembrada de cultivos permanentes y transitorios en el Valle del Cauca.
- Mejorar la planificación agrícola mediante la detección de patrones y la predicción de hectáreas futuras.

Fuentes de Datos

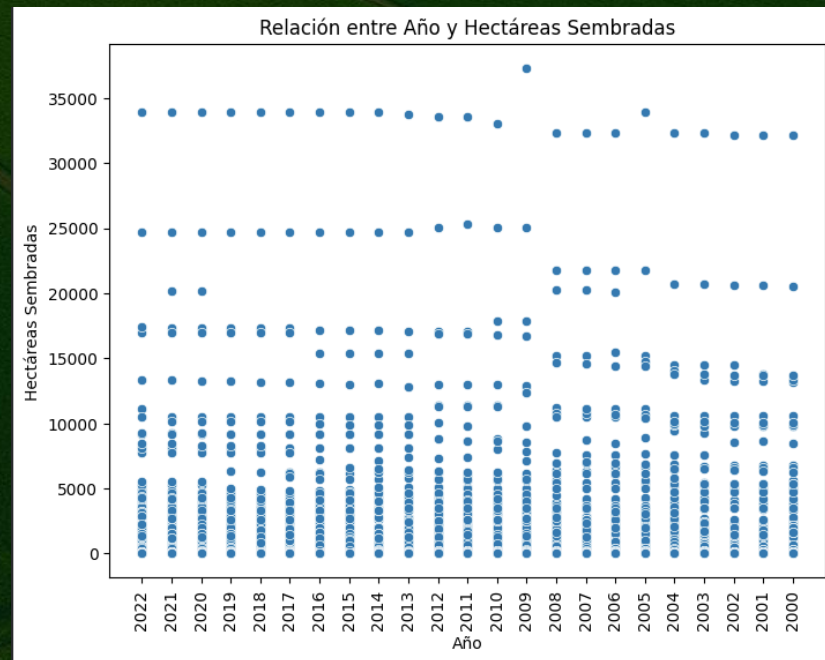
- Fuente: Secretaría de Desarrollo Rural, Agricultura y Pesca
- Portal: datos.gov.co
- Datasets:
 - - Cultivos permanentes
 - - Cultivos transitorios
- Frecuencia: Anual
- Última actualización: Julio 2024

Análisis Exploratorio

- Limpieza y depuración de datos
- Análisis de distribución de cultivos
- Correlaciones entre variables
- Evolución temporal por municipio o tipo de cultivo

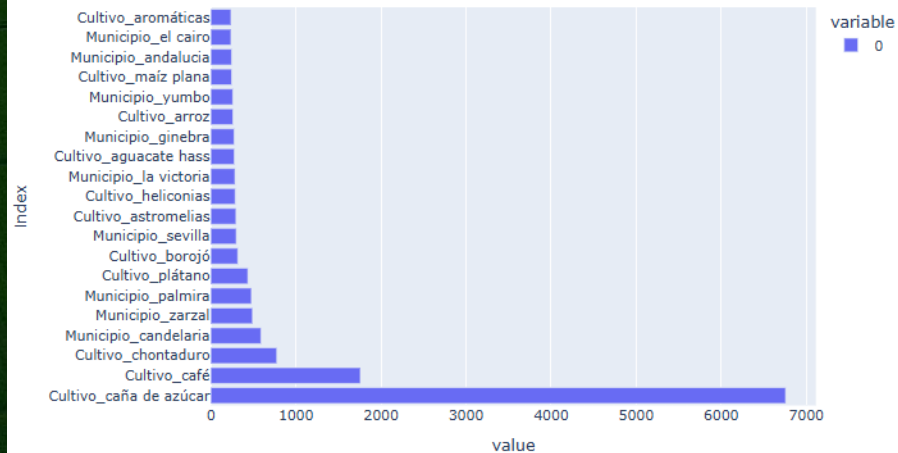
```
RangeIndex: 31338 entries, 0 to 31337  
Data columns (total 6 columns):  
#   Column                Non-Null Count  Dtype  ---  
0   Año                    31338 non-null int64  
1   Municipio              31338 non-null object  
2   Subregión              31338 non-null object  
3   Cultivo                 31338 non-null object  
4   Ciclo                  31338 non-null object  
5   Hectareas_semradas     31338 non-null float64  
dtypes: float64(1), int64(1), object(4)  
memory usage: 1.4+ MB
```

Hectareas_semradas	
count	31338.000000
mean	282.907640
std	1571.553279
min	0.000000
25%	6.000000
50%	18.000000
75%	62.992500
max	37247.400000

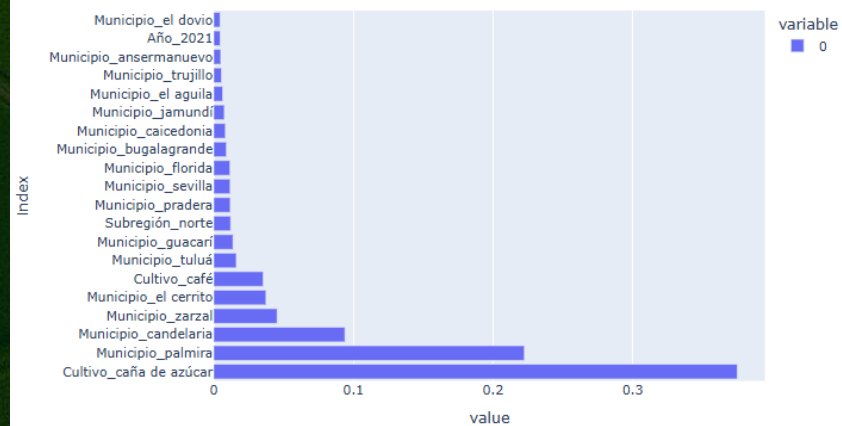


Análisis Exploratorio

Importancia de Variables - Regresión Lineal



Importancia de Variables - Random Forest



Modelos de Machine Learning Aplicados

- Regresión Lineal – relaciones simples año–hectáreas
- Random Forest – relaciones no lineales
- Validación cruzada: 5-fold
- Métricas: MAE, RMSE, R^2
- Importancia de variables: año, municipio, tipo de cultivo

Métricas y gráfico comparativo entre modelos aplicados

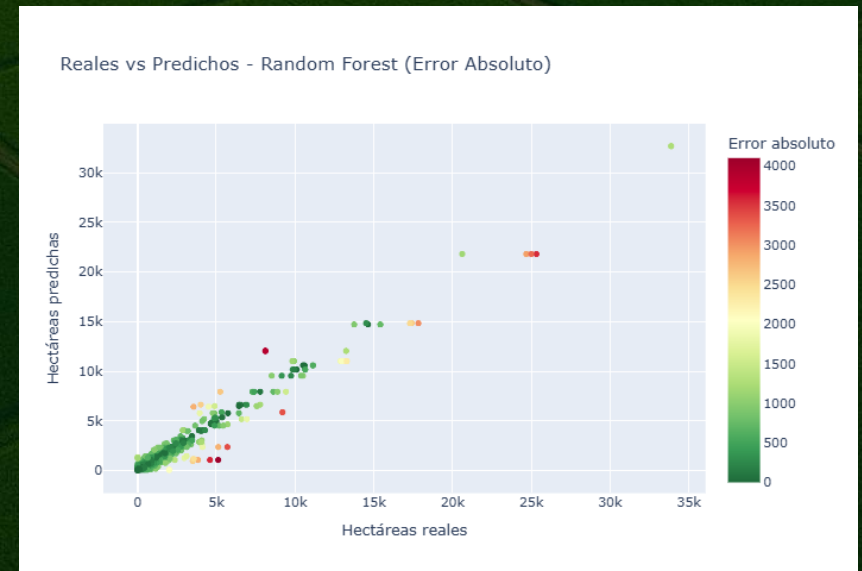
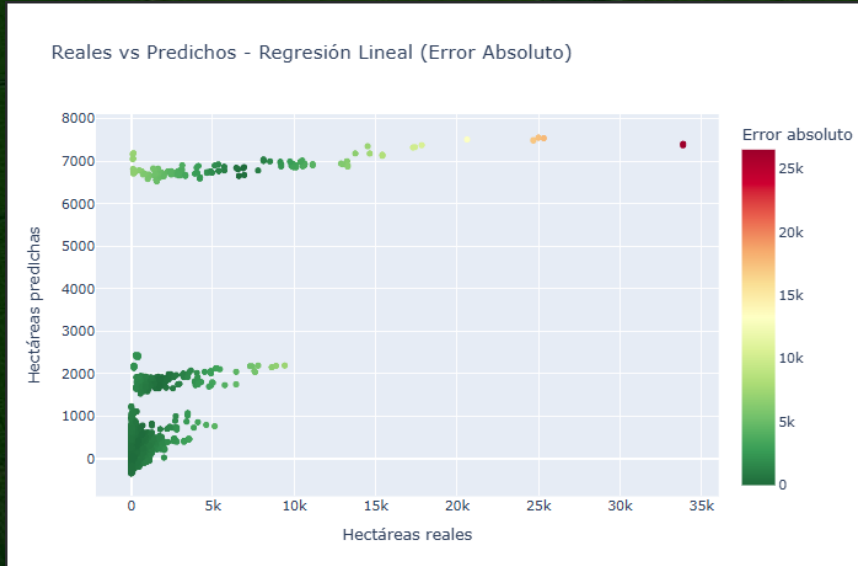


Resultados

- • Random Forest obtuvo el mejor desempeño ($R^2 \approx 0.99$)
- • Regresión lineal útil para interpretación simple
- • MLP mostró sobreajuste (RMSE alto)
- • Modelos predicen con alta precisión las hectáreas futuras
- [Espacio reservado para gráficos comparativos]

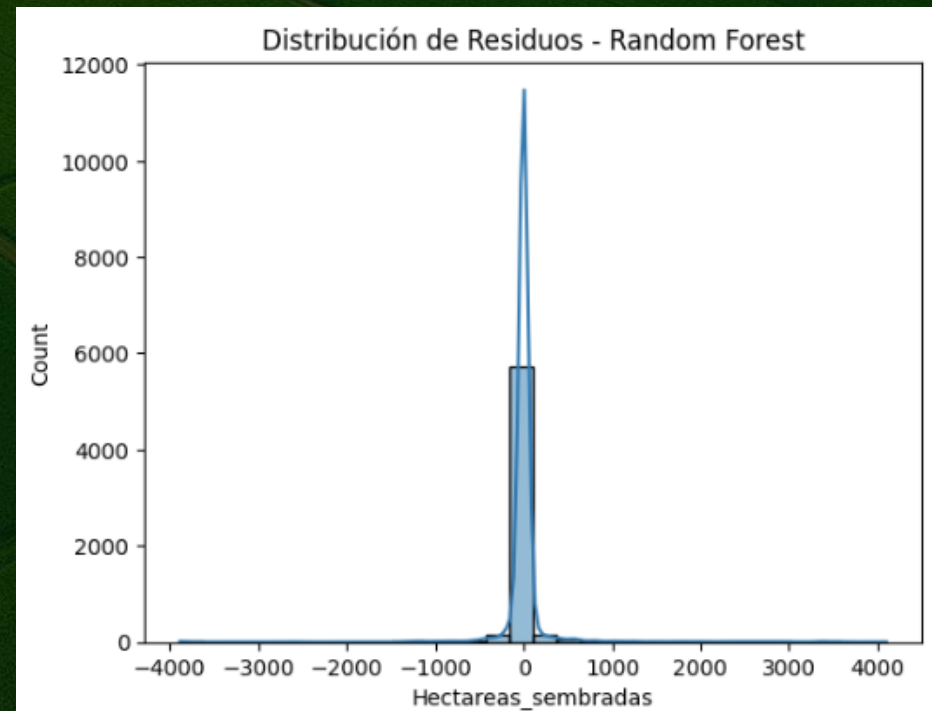
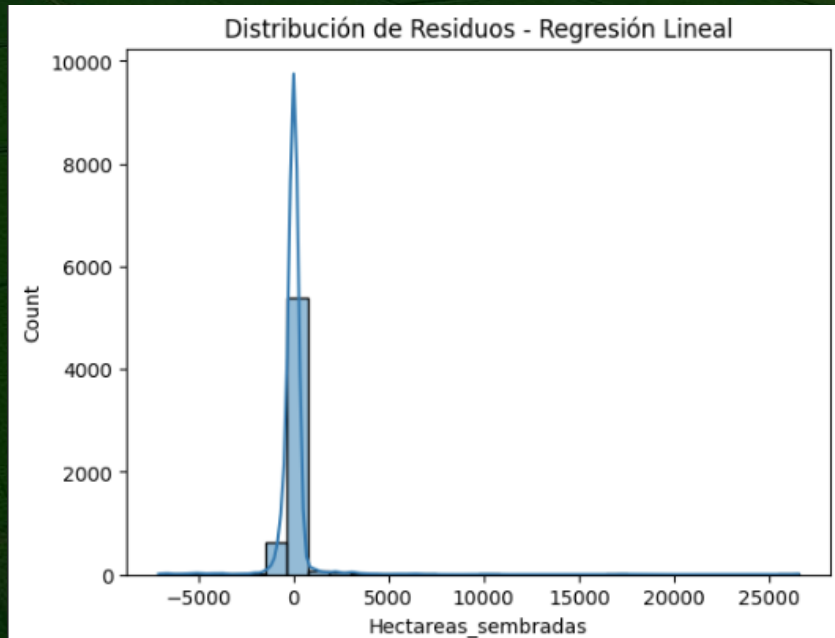
Resultados

- Gráficos de dispersión con error absoluto



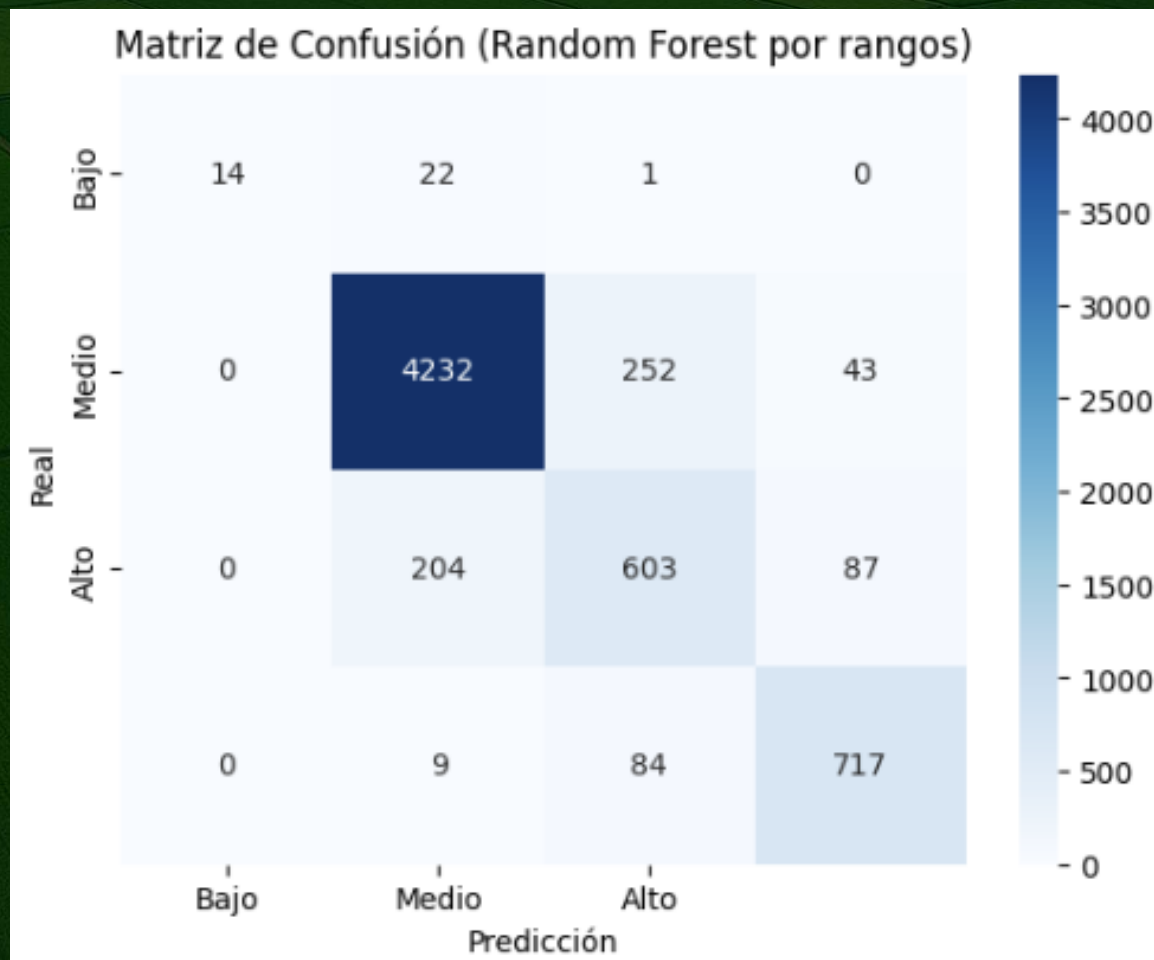
Resultados

- Gráficos de residuos



Resultados

- Matriz de confusión adaptada por rangos



Predicciones

- Se genera 10 ejemplos de predicción:

Ejemplos de predicciones:

	Hectáreas Reales	Predicción Random Forest	Predicción Regresión Lineal	Error RF	Error LR
0	10.0	9.730	-166.244181	0.270	176.244181
1	190.0	156.374	678.683739	33.626	488.683739
2	112.0	85.750	25.830843	26.250	86.169157
3	10.0	22.970	-162.625021	12.970	172.625021
4	77.0	63.030	320.878100	13.970	243.878100
5	30.0	192.550	-145.503998	162.550	175.503998
6	369.0	433.870	197.277456	64.870	171.722544
7	60.0	277.160	-31.259170	217.160	91.259170
8	40.0	189.217	857.224414	149.217	817.224414
9	50.0	9.481	5.531832	40.519	44.468168

Predicciones

- Se diseña un panel para generar predicciones manualmente, el cual si la combinación tiene históricos genera un grafico y compara la predicción de ambos modelos:

<< Opción sin histórico >>

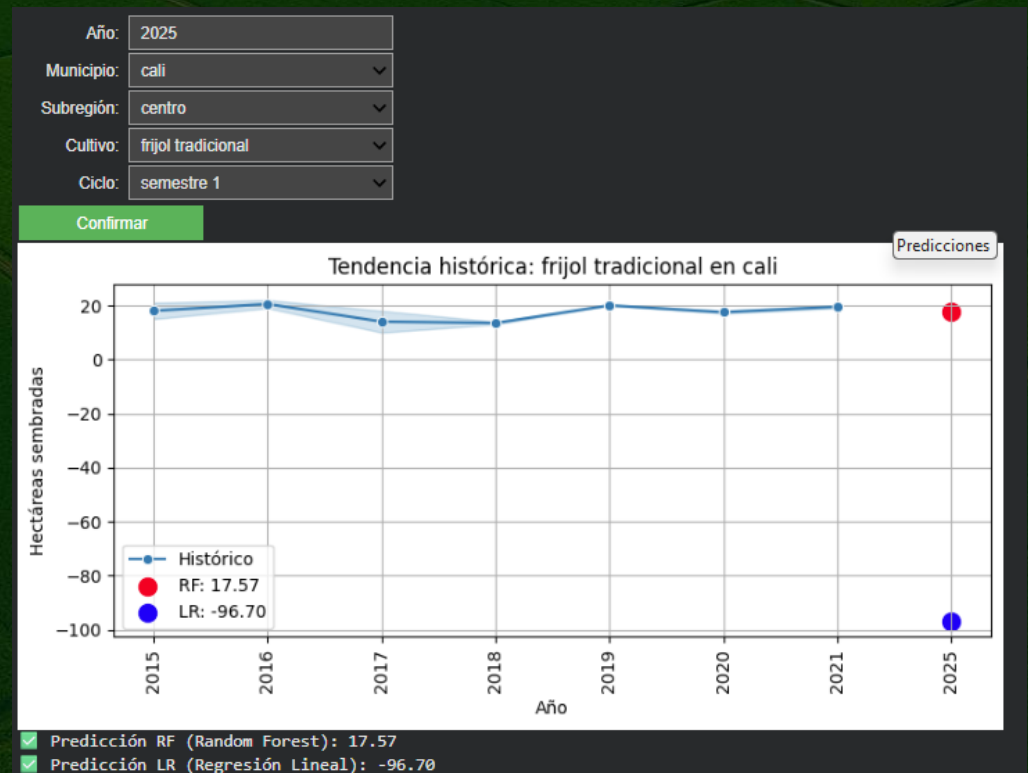
Año:	2025
Municipio:	alcalá
Subregión:	centro
Cultivo:	acelga
Ciclo:	anual

Confirmar

⚠ No hay datos históricos para esta combinación.

- ✓ Se muestran predicciones sin gráfico.
- ✓ Predicción RF (Random Forest): 1.73
- ✓ Predicción LR (Regresión Lineal): 143.78

<< Opción con histórico >>



Aplicaciones del Machine Learning en Agricultura

1. Detección de patrones y tendencias
2. Predicción de áreas sembradas futuras
3. Clasificación por rentabilidad o impacto
4. Integración de cultivos permanentes y transitorios

Conclusiones

- Machine Learning permite generar predicciones precisas y estrategias basadas en datos.
- Mejora la planificación agrícola y el uso sostenible del territorio.
- Los datos abiertos del Gobierno de Colombia son una fuente valiosa para proyectos de IA aplicados al agro.

Referencias

- Portal Nacional de Datos Abiertos – datos.gov.co
- Secretaría de Desarrollo Rural, Agricultura y Pesca
- Notebook: Proyecto Bootcamp IA 2025 – Google Colab

Elaborado por:

- Diego Fernando Velez Neira
- diegofvelez@gmail.com
- Metodología:
 - Metodología Aplicada: CRISP–DM

El presente proyecto de predicción de hectáreas sembradas se desarrolló aplicando la metodología CRISP–DM (Cross Industry Standard Process for Data Mining), un estándar ampliamente utilizado en proyectos de Ciencia de Datos y Machine Learning.

Esta metodología estructura el proceso analítico en seis fases iterativas, asegurando una ejecución ordenada desde la comprensión del problema hasta la obtención de conocimiento útil para la toma de decisiones.