

# **Sistema de Pronóstico de Demanda Diaria**

## **Desafío M5 Forecasting (Walmart)**

### **Análisis, Modelamiento y Validación**

Diego Melián    Carlos Crisosto    Mauricio Pizarro  
Camilo Mesías    Nolberto Rivera

Universidad Técnica Federico Santa María  
Proyectos Estadísticos

1 de diciembre de 2025

# Agenda

- 1 Definición del Problema
- 2 Exploración de Datos (EDA)
- 3 Metodología y Preprocesamiento
- 4 Modelamiento y Resultados
- 5 Interpretabilidad
- 6 Conclusiones y Cierre

# 1. Contexto y Definición del Problema

## Objetivo del Proyecto

Desarrollar un sistema de pronóstico para estimar las ventas diarias unitarias de productos de Walmart por tienda y estado para un horizonte de 28 días.

- **Relevancia:** La optimización de inventario reduce costos logísticos (stock-outs vs. sobre-stock).
- **Alcance (M5 Dataset):**
  - 30,490 series temporales jerárquicas.
  - 3 Estados (CA, TX, WI) y 10 Tiendas.
  - Categorías: Hobbies, Household, Foods.
- **Desafío Técnico:** Series con alta intermitencia (muchos ceros) y alta volatilidad.

## 2. Exploración de Datos: Fuentes

Se integraron tres fuentes de información principales:

- ① **Sales Train/Validation:** Histórico de ventas diarias ( $d\_1$  a  $d\_1913$ ).
- ② **Calendar:** Eventos, feriados, días de la semana y marcadores SNAP (beneficios sociales).
- ③ **Sell Prices:** Precios de venta semanales por tienda e ítem.

### Hallazgo en Calidad de Datos

Se detectó un 68 % de esparsidad (ceros) en promedio.

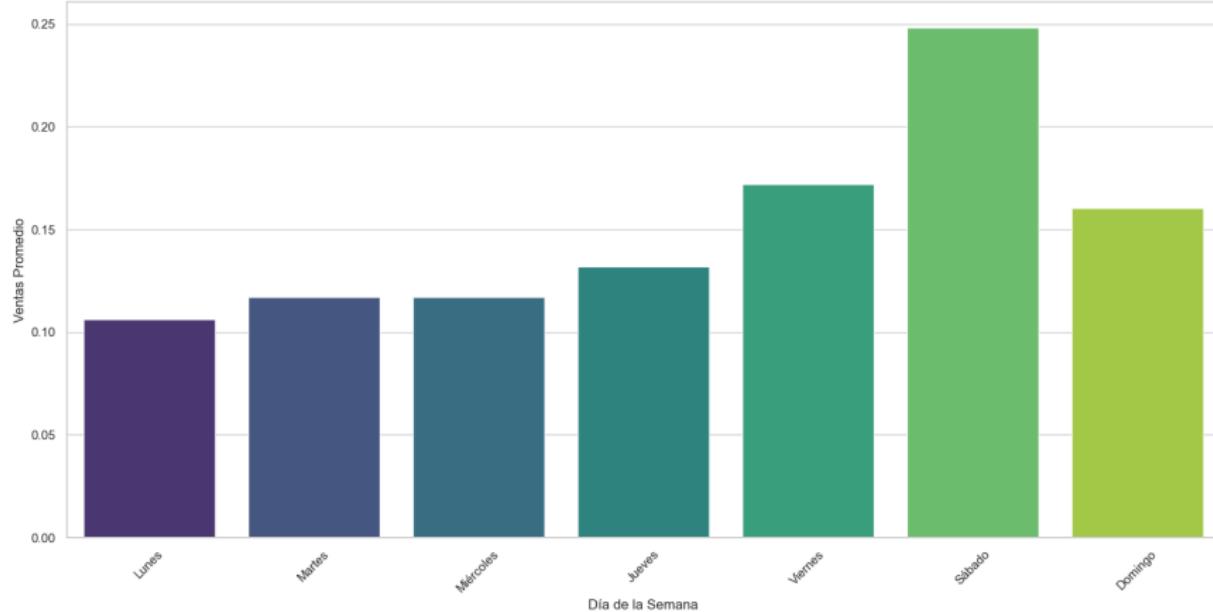
- **Hobbies:** Categoría más intermitente ( 77 % ceros).
- **Foods:** Categoría más estable ( 62 % ceros).

## 2.1 Visualización: Patrón Semanal

### Comportamiento por Día:

- **Peak de ventas:** Viernes y Sábado.
- **Baja demanda:** Inicio de semana.

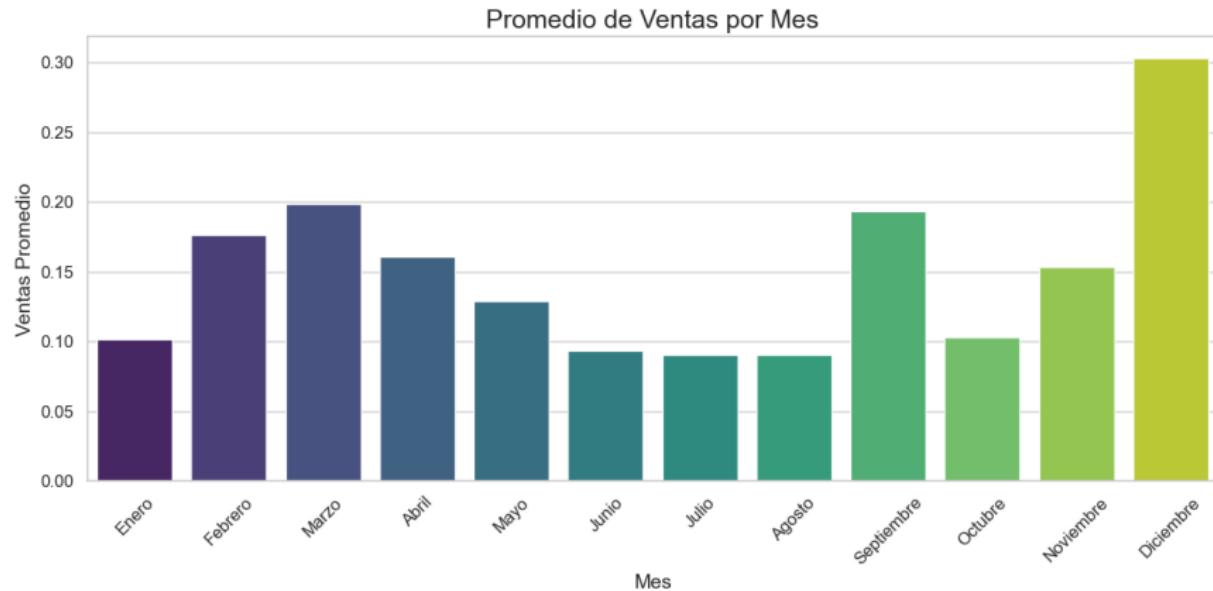
Promedio de Ventas por Día de la Semana



## 2.2 Visualización: Patrón Mensual

### Comportamiento por Mes:

- **Alta demanda:** Diciembre (efecto Navidad/Fin de año).
- **Tendencia:** Creciente observada desde 2014.



### 3. Ingeniería de Características (Feature Engineering)

Para transformar las series temporales en un problema de aprendizaje supervisado, se generaron las siguientes variables:

#### Transformaciones

- **Formato Largo (Melt):** Optimización de memoria usando tipos int16 y float16.
- **Lags (Retardos):** Se generó lag\_28 (ventas de hace 4 semanas) para respetar el horizonte de predicción.
- **Rolling Windows:** Promedios móviles de 7 y 28 días sobre el lag.
- **Calendario:** Día de la semana, mes, año, y eventos SNAP unificados.
- **Precios:** Inclusión del precio de venta semanal.
- **Encoding:** Label Encoding para variables categóricas (Item, Store, Dept).

## 4. Estrategia de Validación

Se utilizó una validación cruzada tipo **Walk-Forward** (ventana deslizante) para respetar la causalidad temporal.

- **Horizonte:** 28 días por split.
- **Splits:** 3 validaciones consecutivas (últimos 3 meses del dataset).
- **Métricas de Evaluación:**
  - **RMSE:** Error cuadrático medio (sensible a outliers).
  - **MAE:** Error absoluto medio (interpretabilidad).
  - **RMSSE:** Métrica oficial de M5 (Error escalado, robusto a diferentes escalas de venta).

## 5. Modelos Evaluados

Se implementó un enfoque comparativo con diferentes familias de modelos:

### ① Baselines (Línea Base):

- *Naive*: Repite el último valor.
- *Seasonal Naive (SNaive)*: Repite el valor de hace 7 días.

### ② Modelos Lineales:

- *Regresión Lineal*: Utilizando lags y features de calendario.

### ③ Machine Learning (Gradient Boosting):

- *LightGBM*: Modelo basado en árboles, optimizado para velocidad y manejo de grandes volúmenes de datos.

### ④ Series de Tiempo (Agregadas):

- *SARIMA y Prophet*: Evaluados sobre la venta total diaria para capturar tendencia global.

## 6. Comparación de Modelos – RMSE en Validación Walk-Forward

Modelo	Split 1 (RMSE)	Split 2 (RMSE)	Split 3 (RMSE)
Naive (Baseline)	3.13	3.14	3.30
SNaive (Estacional)	3.67	3.63	3.60
Regresión Lineal	2.45	2.35	2.24
<b>LightGBM</b>	<b>2.78</b>	<b>2.20</b>	<b>2.13</b>

**Cuadro 1:** Comparación de desempeño (RMSE) en los 3 splits de validación.

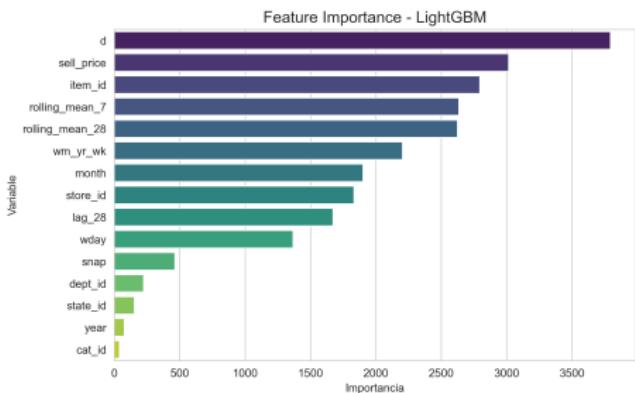
### Análisis

- **LightGBM** superó consistentemente a los baselines y a la regresión lineal en los últimos splits.
- La regresión lineal mostró un desempeño sorprendentemente robusto, indicando que la tendencia y los lags son predictores fuertes.
- SNaive tuvo el peor desempeño, sugiriendo que la demanda no es puramente cíclica semanal.

## 7. Interpretabilidad del Modelo (LightGBM)

### Importancia de Variables:

- ① **d**: Indicador temporal más influyente.
- ② **Sell Price**: Fuerte relación con el nivel de ventas.
- ③ **Item ID**: Comportamiento propio del producto.
- ④ **Rolling Mean (7, 28)**: Señal de tendencia reciente.
- ⑤ **Calendario**: Efectos de días de la semana y meses.



**Insight Accionable:** Las estrategias de precios y la gestión de inventario basada en la media móvil reciente son críticas.

## 8. Conclusión Ejecutiva

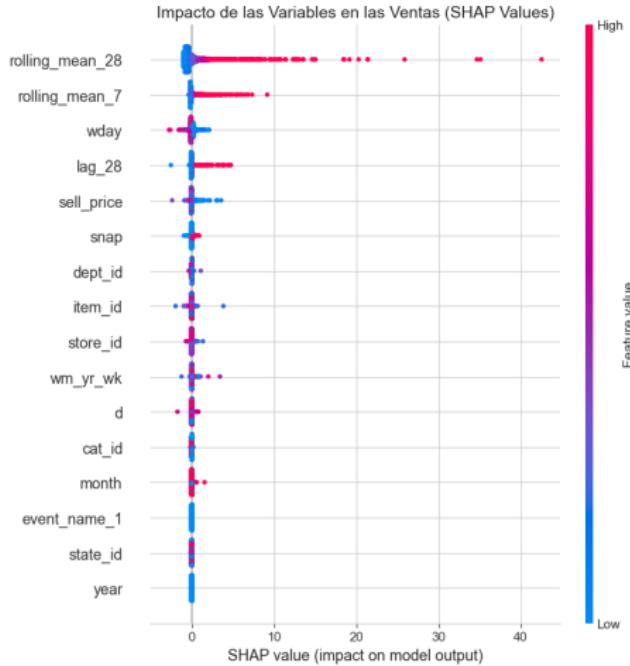
### Modelo Recomendado: LightGBM

Se selecciona LightGBM por ofrecer el mejor equilibrio entre precisión y eficiencia computacional.

### Hallazgos Clave:

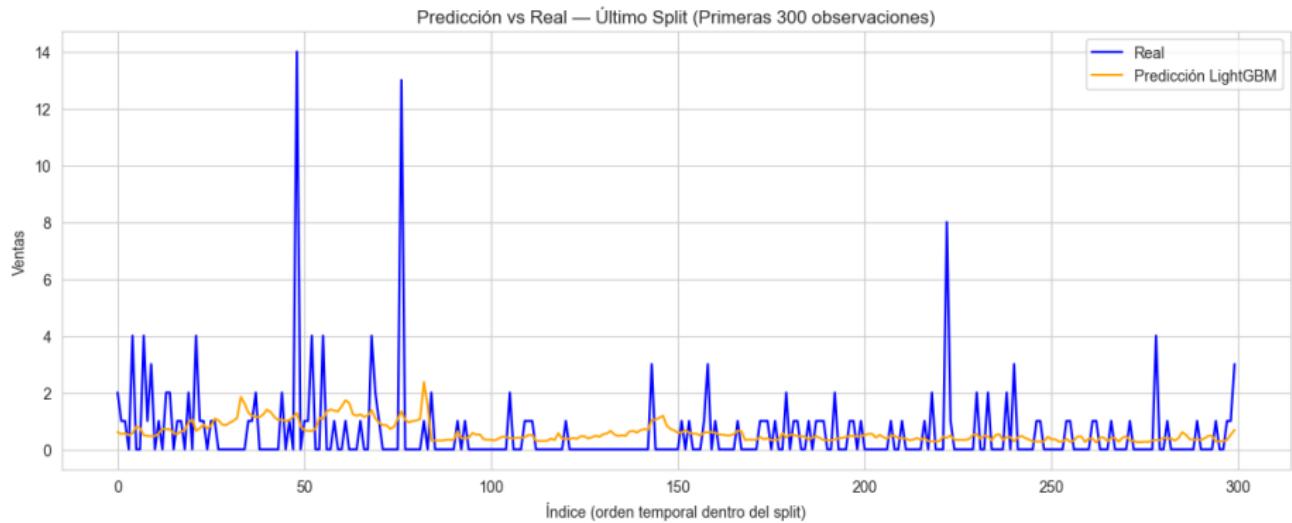
- **Valor Práctico:** El modelo logra una reducción del error cercana al **30 %** respecto al baseline Naive.
- **Comportamiento por Segmento:** El error es considerablemente menor en la "Long Tail" (productos de venta esporádica, MAE 0.70) en comparación con los *Top Sellers* (MAE 2.54). Esto es ideal para una empresa grande como Walmart, donde la mayoría del catálogo son productos de baja rotación.

## 8.1 Drivers de Venta (Interpretabilidad)



Los valores SHAP indican que el precio y la inercia reciente (*rolling mean*) son los predictores más fuertes.

## 8.1 Conclusión Ejecutiva



**Figura 1:** Relación entre valores predichos y reales en el último split de validación.

## 8.2 Auditoría de Desempeño por Subgrupos

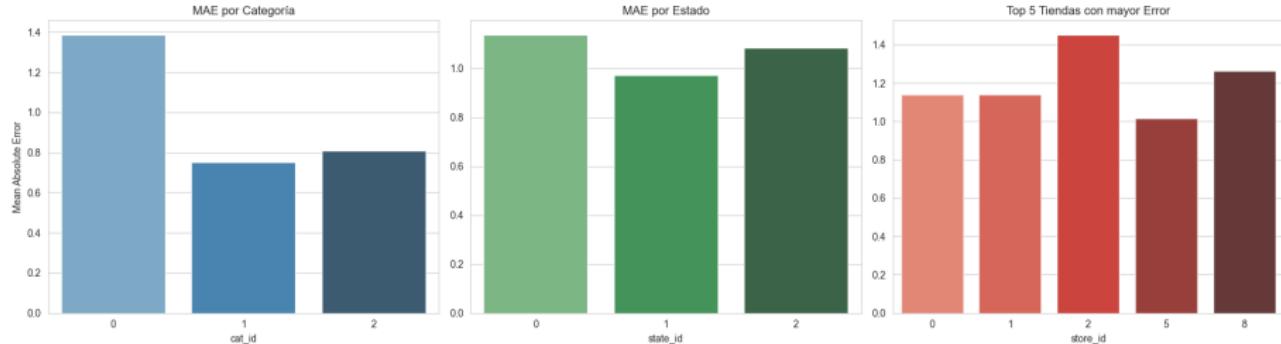
### Objetivo del Análisis

Para garantizar la robustez operativa, se desglozó el error (MAE) en tres dimensiones clave: Categoría, Estado y Tienda.

### Principales Observaciones:

- **Por Categoría:** Generalmente, las categorías de *Foods* presentan mayores errores , mientras que *Hobbies* es más difícil de predecir en términos relativos por su intermitencia.
- **Por Geografía (Estado):** Este análisis permite detectar si factores locales afectan la calidad de la predicción.
- **Foco en Tiendas Críticas:** Identificar el "Top 5" de tiendas con mayor error permite dirigir esfuerzos de re-entrenamiento específico para esos locales en particular.

## 8.3 Detalle Visual: Errores por Segmento



*Desglose del Error Absoluto Medio (MAE) a través de las jerarquías del negocio.*

## 9. Limitaciones y Riesgos

- **Limitación de Memoria:** Debido al gran volumen (30k series x 1900 días), se debió trabajar con muestras y tipos de datos optimizados, lo que podría ocultar patrones en series muy específicas.
- **Datos Externos:** No se incluyeron datos de clima o promociones de la competencia, que podrían explicar varianza residual.
- **Horizonte Fijo:** El modelo pierde precisión más allá de los 28 días si no se actualizan los lags (requiere re-entrenamiento o enfoque recursivo).

## 10. Próximos Pasos / Llamada a la Acción

- ① **Implementación Recursiva:** Configurar el pipeline para predecir día a día actualizando los lags dinámicamente para un entorno de producción real.
- ② **Deep Learning:** Explorar modelos como *DeepAR* o *Temporal Fusion Transformers* para capturar dependencias globales complejas, evaluando si la mejora justifica el costo computacional.
- ③ **Ensamble:** Combinar las predicciones de LightGBM con Prophet para robustecer la captura de tendencias a largo plazo y estacionalidades anuales.

# ¡Gracias por su atención!

Universidad Técnica Federico Santa María  
Proyectos Estadísticos