

# ML PARA OPTIMIZAR PRONÓSTICOS DE VENTAS DE REFRESCOS CON SEGMENTACIÓN Y VARIABLES CLIMÁTICAS

Marcos Aquino, Natalia Olvera, Ericka Rodríguez, Carol Rendón & Rolando Ruiz

## TC3007C Inteligencia Artificial Avanzada para la Ciencia de Datos II

### Resumen

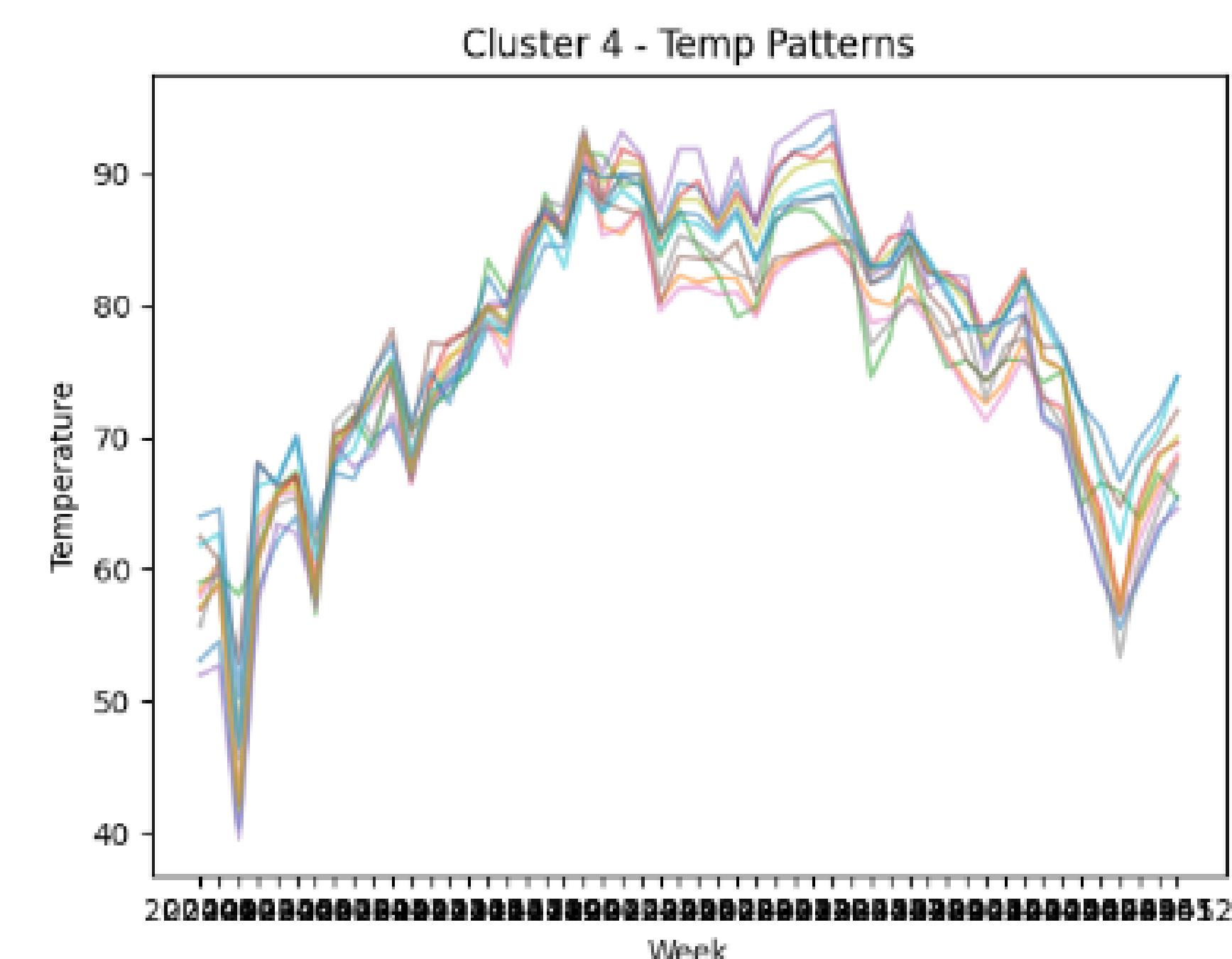
El proyecto busca mejorar la predicción de ventas de refrescos en canales tradicionales (abarros, kioskos y tienditas) mediante modelos de aprendizaje máquina que incorporan variables climáticas como la temperatura. Se utilizan datos históricos de ventas, meteorológicos y las predicciones originales de la empresa. La metodología incluye análisis exploratorio, agrupamiento de regiones con clima similar, construcción de un modelo base y modelos avanzados (regresión lineal y XGBoost), evaluados con MAE. La iniciativa permite optimizar inventarios y distribución, reduciendo exceso y desabasto de productos.

### Introducción

En la industria de bebidas, la precisión en los pronósticos de venta es clave para la eficiencia y rentabilidad, ya que errores por desabastecimiento o sobrestock afectan producción, distribución y satisfacción del cliente. La cadena de suministro, desde la embotelladora hasta el punto de venta, depende de anticipar la demanda con exactitud. Factores como clima, precio, promociones y hábitos de consumo influyen en la demanda, pero muchos modelos actuales solo consideran series históricas y variables comerciales. Este estudio busca mejorar la predicción de ventas B2B incorporando variables climáticas, evaluando cómo la inclusión de datos meteorológicos locales puede optimizar los modelos tradicionales y aumentar la precisión semanal por categoría de producto a nivel de punto de venta.

### Metodología

- Clusterización climática: Agrupación de 33 estaciones meteorológicas según patrones semanales.
- Ingeniería de datos: Creación de estadísticas semanales y dinámicas de temperatura para alinear con la granularidad de ventas.
- Modelos de pronóstico: Construcción de un modelo base, regresión lineal y XGBoost para evaluar la incorporación de variables climáticas.
- Evaluación: Comparación del desempeño de los modelos usando MAE y análisis de la mejora obtenida comparado con el modelo inicial.



### Modelos Realizados

Se evaluaron los siguientes modelos :

1. Modelo Base: Creamos un factor de ajuste en base a reglas preestablecidas con el clima a nivel de clúster y producto.
2. Regresión Lineal: La regresión lineal multivariada se utilizó como primer modelo paramétrico para estimar la demanda semanal. Este modelo asume una relación lineal entre la variable objetivo (ventas semanales por cliente) y el conjunto de predictores utilizados
3. XGboost: El modelo XGBoost se empleó como alternativa no lineal y más expresiva para capturar relaciones complejas entre clima, ventas pasadas y patrones de interacción cliente-semana.

### Conclusión

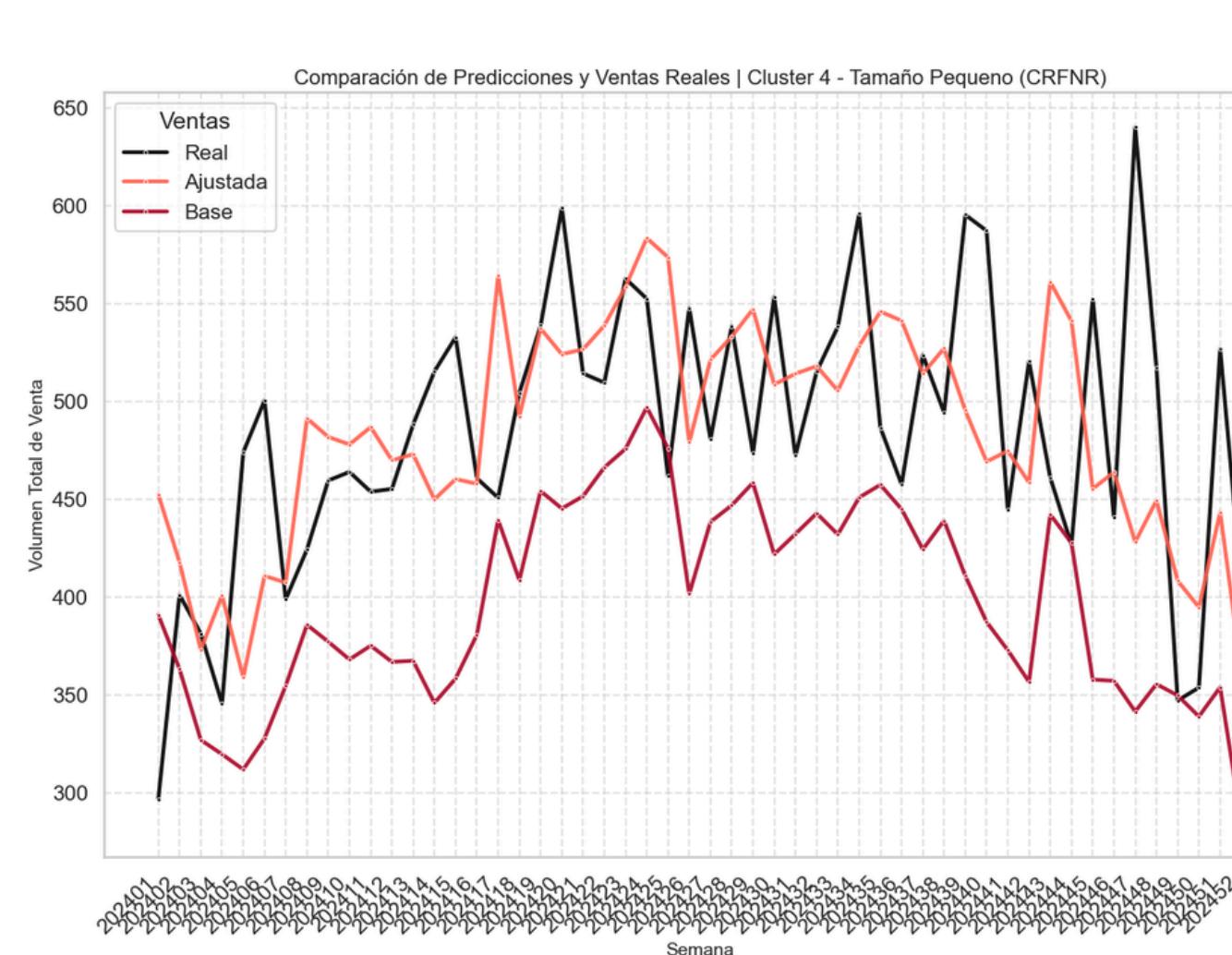
Aunque se evaluaron múltiples modelos, los resultados obtenidos no permiten confirmar la hipótesis de que las variables climáticas mejoren el pronóstico de ventas de refrescos en México. En esta etapa del análisis, la inclusión de dichas variables incrementó el error de predicción. Sin embargo, estos hallazgos sugieren la necesidad de explorar ajustes en la selección de variables, niveles de granularidad, modelos más sofisticados o períodos adicionales de datos, lo que abre oportunidades para investigaciones posteriores.

### Resultados

#### Modelo Base

Error original	Error Modelo
18.3	19.7

Producto	Error Original	Error Modelo	Mejora
CRFNR	7.24	7.48	-3.30
CRFR	7.48	7.55	-0.93



#### Modelos ML

Modelo	Error base	Error
Regresión Lineal	7.62	8.05
XGBoost	7.62	8.51

Producto	Modelo	Error base	Error modelo	Mejora
CRFNR	Lineal	7.54	8.00	-0.47
CRFNR	XGBoost	7.53	8.46	-0.93
CRFR	Lineal	7.69	8.09	-0.40
CRFR	XGBoost	7.70	8.56	-0.86

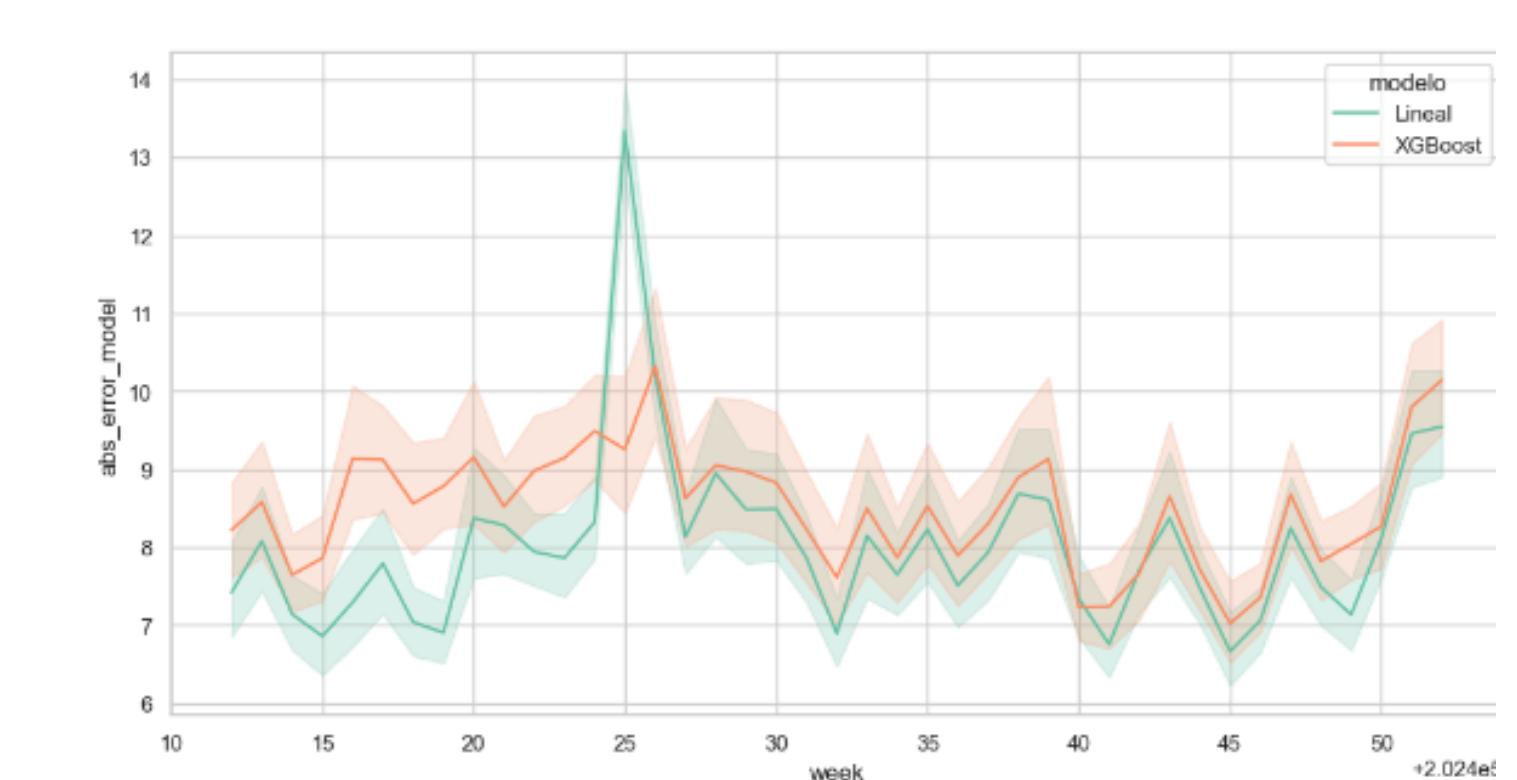


Figura 2: Error de modelo por semana