

# MODELADO DE PRECIOS EN LA CANASTA BÁSICA USANDO MACHINE LEARNING

HERNÁNDEZ IBARRA Gladys Guadalupe

gladys.hernandezibr@uanl.edu.mx, Maestría en Ciencia de Datos

## Introducción

Este estudio analiza la base de datos de la PROFECO, que incluye los precios semanales de productos supervisados por la institución desde 2015 hasta 2023<sup>3</sup>. El foco de atención está en detectar aumentos injustificados de precios, especialmente en productos como huevo, tortilla, leche, frijol y carne de res<sup>2</sup>. El objetivo principal es explorar técnicas de machine learning para modelar los precios de los productos básicos.

## Metodología

- Fechas: se considero la base desde 2015 a 2022, se tomó solamente la primera semana de cada mes.
- Se considero información de toda la República, para análisis donde se agrupan los datos se considero por separado zonas rurales y urbanas<sup>1</sup>.
- Solamente se consideran los productos de el Huevo, la Tortilla, Leche, Frijol, y Carne de Res.
- Se consideran las variables Presentación, Marca, Cadena Comercial, Giro, Estado, año, mes, y día.

Tabla.1. Estadísticos de la variable precio por producto

Estadístico	Mean	STD	MIN	25%	50%	75%	MAX
Carne	142.9	71.5	17.0	96.0	139.0	164.0	659.0
Frijol	31.4	10.8	8.5	23.5	29.9	37.9	73.9
Huevo	36.9	14.9	13.4	26.5	33.5	42.5	106.0
Leche	18.1	5.3	5.0	15.3	17.6	21.4	36.9
Tortilla	14.1	3.7	7.0	11.0	13.5	16.0	32.0

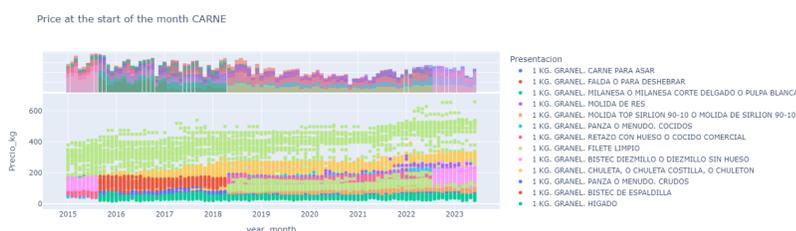


Figura.1. Precios de la carne seccionado por presentación

### Diseño del experimento y modelos

- Se implemento el uso de GridSearchCV para ajustar los parámetros donde fuera apropiado, así como KElbowVisualizer para KNN-R
- Para variables categóricas se implemento one hot encoder y para variables numéricas StandardScaler

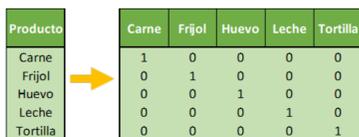


Figura.2. Onehot-encoder Ilustrativo

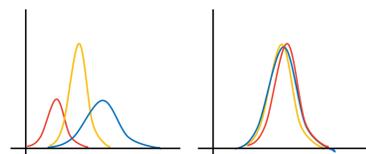


Figura.3. StandardScaler Ilustrativo

### Modelos

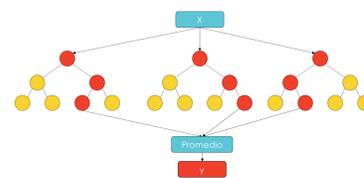


Figura.4. Ilustración del proceso de Random Forest

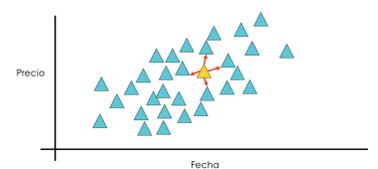


Figura.5. Ilustración del proceso de KNN

### Métricas de rendimiento

- MAE: Por su facilidad de interpretación.
- RMSE: Complementa el MAE penalizando posibles errores más grandes.
- MAPE: Permite comprar los errores sin depender de la escala de los datos.

## Resultados

Caso de Carne de Res 1 KG BISTEC en Nuevo León



Figura.6. Gráfica mostrando los valores reales del precio de la carne vs los modelados para un caso específico en Nuevo León

Tabla.2. Tabla con las métricas de rendimiento por modelo y por horizonte de tiempo en el entrenamiento

Modelo	Métricas de Rendimiento	2015-2022	2020-2022	2021-2022
RF-R	MAE	23.160	20.157	17.750
	RMSE	46.194	39.508	34.357
	MAPE	0.323	0.369	0.337
KNN-R	MAE	18.812	10.742	8.018
	RMSE	27.199	17.286	14.500
	MAPE	0.299	0.175	0.126

## Conclusiones

Este estudio empleó técnicas de modelado de machine learning para explorar la variación en el rendimiento en función de su horizonte temporal y su capacidad para capturar eventos socioeconómicos cambiantes. Se identificó que el algoritmo KNN-R exhibe un rendimiento destacado en la explicación de datos, lo que lo posiciona como una herramienta útil para detectar potenciales desviaciones inesperadas en los precios y sugiere su potencial como una alerta temprana para investigar dichos casos.

## Referencias

[1] S. de Información Agroalimentaria y Pesquera. ¿qué nos dice el censo de población sobre la juventud rural de nuestro país?, 2022. URL <https://www.gob.mx/siap/articulos/que-nos-dice-el-censo-de-poblacion-sobre-la-juventud-rural-de-nuestro-pais>.

[2] G. de Puebla. Por ley profeco debe vigilar y regular los precios, 2019. URL <https://puebla.gob.mx/index.php/noticias/item/1426-por-ley-profeco-debe-vigilar-y-regular-los-precios>.

[3] PROFECO. Quien es quien en los precios, 2023. URL [https://datos.profeco.gob.mx/datos\\_abiertos/](https://datos.profeco.gob.mx/datos_abiertos/).