

# Web Scraping y modelos de forecasting aplicados a precios de subastas ganaderas en colombia

*Anónimo*

**Abstract** Con el objetivo de entender el comportamiento de los precios base que se aplican en la subasta CENCOGAN, se llevó a cabo un proceso de extracción de datos utilizando la técnica de Web Scraping en las páginas oficiales de CENCOGAN, Fedegan (Federación Nacional de Ganaderos) y datos abiertos del gobierno colombiano. Luego, la información recolectada fue analizada en detalle mediante el uso del programa RStudio, aplicando una serie de técnicas de análisis estadístico descriptivo, exploratorio univariado, bivariado y multivariado. Una vez completado el análisis descriptivo, se aplicaron técnicas de estadística inferencial y de pronóstico (forecasting) para modelar la variación de los precios a lo largo del tiempo. En este sentido, se utilizaron tanto modelos de regresión lineal múltiple como modelos Autorregresivos (ARIMA), con el fin de identificar patrones y tendencias a lo largo del tiempo. En conclusión, este estudio proporciona una visión completa y detallada del comportamiento de los precios base en la subasta CENCOGAN, utilizando una variedad de técnicas de análisis estadístico y herramientas de modelado predictivo. Los resultados obtenidos pueden ser de gran utilidad para los inversores y demás actores del mercado ganadero, permitiéndoles tomar decisiones más informadas y precisas en cuanto a la gestión de sus operaciones y estrategias comerciales.

**Palabras clave:** Web Scraping - Series Temporales - Subastas Ganaderas

## Antecedentes

Para comprender los sucesos aleatorios de fenómenos que no se entienden, se usan diversos métodos estadísticos, estos también se emplean en las subastas de ganado bovino, ya que, los cambios en los precios y la venta afectan directamente a los ingresos de las explotaciones bovinas. Algunos ejemplos de estos métodos son los modelos aditivos lineales, la media móvil multiplicativa, la tasa de crecimiento sobre las medias anuales, los modelos condicionales heterocedásticos autorregresivos y el modelo lineal econométrico auto-regresivo con retrasos distribuidos.

## Objetivos

### Objetivo General

Analizar la variabilidad temporal de los precios base que se aplican en la subasta CENCOGAN mediante el uso de técnicas de modelado predictivo.

### Objetivos específicos.

- Automatizar el flujo operacional de Web Scraping para extracción de fuentes de información.
- Comparar modelos de forecasting para predicción de precios de subastas.

## Metodología

La extracción de datos se realizó mediante la técnica de Web Scraping en las páginas oficiales de CENCOGAN, Fedegan y Datos Abiertos del gobierno colombiano. La información recolectada fue analizada mediante el software R con RStudio, aplicando técnicas de análisis estadístico descriptivo, exploratorio univariado, bivariado y multivariado. Se aplicaron técnicas de pronóstico para modelar la variación de los precios a lo largo del tiempo. Se utilizaron modelos de regresión lineal múltiple, modelos Autorregresivos (ARIMA) y aproximaciones de machine learning para series temporales. Los modelos fueron comparados con la métrica Error Absoluto Porcentual Medio (MAPE) y la Raíz del Error Cuadrático Medio (RMSE). Todos los análisis fueron desarrollados con el software estadístico R (R Core Team) a través del Entorno de Desarrollo Integrado (IDE) Rstudio y Bibliotecas especializadas en diferentes procedimientos de acceso, transformación, visualización y modelamiento de datos fueron utilizadas para extracción de información, análisis descriptivo y exploratorio y comparación de modelos estadísticos. Algunas de las bibliotecas de uso frecuente en este trabajo fueron: rvest, tidyverse, FactoMineR, tidymodels, modeltime.

## Resultados

Con los modelos evaluados se obtuvo un MAPE de 9.15 % hasta 11.89 % y de 4.28 % hasta 22.42 %, para predicciones globales y locales, respectivamente.

Se evidenció que la volatilidad en el precio base del ganado incrementó después del COVID19, también fue posible dilucidar que categorías como el toro y hembras de vientre exhiben comportamientos más homogéneos en el precio.

Se encontraron correlaciones altas y positivas entre variables exógenas como la tasa representativa del mercado y el precio base del ganado.

## Conclusiones

- El uso de técnicas de análisis estadístico y herramientas de modelado predictivo son de alta relevancia para comprender la variabilidad temporal del precio base del ganado, constituyéndose en un elemento de utilidad para los inversores y demás actores del mercado ganadero, permitiéndoles tomar decisiones anticipadas y precisas en cuanto a la gestión de sus operaciones y estrategias comerciales.
- La implementación de técnicas de web scraping para la automatización de extracción de información permitió acelerar el proceso de análisis mediante el uso de datos abiertos del gobierno colombiano.