

Informe Final de Interpretación de Resultados y Recomendaciones

Autor: Kristhian Santiago Palomino Fajardo

Fecha: 18 de marzo de 2025

Introducción

El objetivo de este informe es interpretar los resultados del modelo predictivo desarrollado para estimar el número de unidades vendidas (Units_Sold) a partir de datos históricos. Para ello, se han analizado exhaustivamente las características del dataset, se han aplicado procesos de preprocesamiento robustos y se han entrenado varios modelos de regresión (lineales y ensemble). Este informe final ofrece una interpretación de los coeficientes, identifica posibles mejoras en la precisión del modelo y brinda recomendaciones estratégicas para el equipo de ventas.

Interpretación de los Resultados del Modelo Predictivo

Este análisis permite identificar de forma detallada el impacto de cada característica en la predicción de Units_Sold, evidenciando la importancia relativa de cada variable en el comportamiento de las ventas.

- **Variables relacionadas con Store y Category:**
Se observa que las dummies correspondientes a “Store_102” y “Category_Clothing” tienen coeficientes positivos significativos, lo que indica que estas variables están asociadas a un aumento en las ventas. Por el contrario, las variables “Store_103” y “Category_Home Goods” presentan coeficientes negativos, sugiriendo una relación inversa, consistente con la diferencia en rangos de precios y comportamiento de ventas.
- **Variables temporales (Día_de_la_semana):**
Las diferentes dummies de los días de la semana muestran variaciones en los coeficientes. Por ejemplo, los días centrales (martes y miércoles) tienden a tener coeficientes que impulsan las ventas, mientras que otros días pueden tener un impacto menor o incluso negativo. Esto respalda la observación del EDA, en donde se identificó mayor actividad en los días centrales de la semana.
- **Unit_Price y Day_of_Month:**
La estandarización de estas variables permite que sus coeficientes sean comparables. En los modelos lineales, se observa que Unit_Price tiene un coeficiente negativo moderado, sugiriendo que productos con precios más elevados, en promedio, se asocian a menores volúmenes de venta, lo cual es coherente con la naturaleza de las categorías (Electronics con precios altos vs. Clothing y Home Goods con precios bajos).

Evaluación de Modelos

Modelo	MAE	RMSE	R ²
Linear Regression	4.90	5.77	0.80
Lasso	4.89	5.83	0.80
Ridge	4.79	5.73	0.81
Decision Tree	9.20	10.70	0.32
Random Forest	5.92	7.27	0.69
Gradient Boosting	6.65	8.23	0.60
AdaBoost	4.77	6.56	0.74

Interpretación:

Los modelos lineales (especialmente Ridge) explican aproximadamente el 81% de la varianza en las ventas, lo que indica un desempeño sólido con buena capacidad explicativa. Su interpretabilidad permite extraer insights claros sobre el impacto de cada variable. Aunque algunos métodos ensemble como AdaBoost muestran un MAE ligeramente inferior, el desempeño global (especialmente R²) de los modelos lineales es superior. Esto se debe, en parte, a la naturaleza del dataset en el que cada tienda vende una única categoría, limitando la variabilidad que pueden capturar los métodos más complejos. Este modelo muestra un desempeño significativamente inferior, lo que sugiere una alta sensibilidad al sobreajuste y una baja robustez en el contexto de estos datos.

Identificación de Posibles Mejoras en la Precisión del Modelo

Para mejorar aún más la precisión del modelo predictivo se recomienda:

- Recopilar datos a lo largo de varios meses o años permitirá capturar patrones estacionales y tendencias a largo plazo. Esto no solo ayudará a ajustar mejor los modelos sino que también permitirá evaluar la robustez de las predicciones frente a variaciones temporales.
- La integración de información sobre promociones, eventos especiales, días festivos y condiciones económicas podría explicar variaciones en las ventas que actualmente no se consideran. Estas variables adicionales permitirían al modelo capturar factores contextuales y mejorar su capacidad predictiva.

Recomendaciones Prácticas para el Equipo de Ventas

Con base en los resultados y la interpretación del modelo, se sugieren las siguientes recomendaciones estratégicas:

Estrategia Diferenciada por Categoría:

Dado que cada tienda se especializa en una única categoría, es crucial adaptar las estrategias comerciales:

- Electronics (Tienda 101): Focalizar promociones de alto valor y asegurar un stock adecuado, dada la alta contribución en ingresos, a pesar de un menor volumen de unidades vendidas.
- Clothing (Tienda 102): Diseñar campañas que incentiven la venta en mayor volumen, por ejemplo, mediante ofertas de cross-selling o bundles, para incrementar el ticket promedio.
- Home Goods (Tienda 103): Implementar estrategias de marketing o diversificación de productos para aumentar tanto el volumen de ventas como los ingresos totales.

Optimización según Días de la Semana:

Se observó que los días centrales de la semana (martes y miércoles) tienen mayor actividad. Se recomienda:

- Refuerzo en días laborables: Aumentar la dotación de personal y la disponibilidad de inventario en estos días, aprovechando la alta demanda.
- Promociones específicas para fines de semana: Diseñar campañas dirigidas a incentivar las compras durante los días con menor actividad, como los fines de semana.
- Monitoreo Continuo y Retroalimentación: Implementar un sistema de seguimiento en tiempo real del desempeño del modelo y de las ventas. Esto permitirá realizar ajustes ágiles en las estrategias de ventas, garantizando que el modelo se mantenga relevante ante cambios en el comportamiento del consumidor.

Reflexión sobre las Limitaciones del Análisis y Sugerencias para Futuros Trabajos

Limitaciones del Análisis Actual:

- Horizonte Temporal Reducido: El análisis se basa en datos de un único mes, lo que puede limitar la capacidad para identificar patrones estacionales y tendencias a largo plazo. Esto reduce la generalización del modelo en escenarios diferentes o en períodos prolongados.
- Falta de Variables Externas: La ausencia de información sobre promociones, eventos, días festivos y factores económicos impide capturar algunas de las variaciones en las ventas. Estas variables pueden ser fundamentales para comprender comportamientos de compra en contextos específicos.
- Distribución Unívoca de Categorías por Tienda: El hecho de que cada tienda venda una única categoría simplifica el análisis, pero también limita la capacidad del modelo para generalizar estrategias a tiendas con portafolios diversificados.

Sugerencias para Futuros Trabajos:

- Recopilar datos de períodos más largos (varios meses o años) permitirá una mayor robustez en el análisis, facilitando la detección de patrones estacionales y tendencias que no se evidencian en un corto período.
- La incorporación de datos adicionales (promociones, eventos, indicadores económicos) enriquecerá el análisis y permitirá ajustar el modelo a factores contextuales que impactan las ventas
- Desarrollar un dashboard o herramienta de monitoreo que permita evaluar el desempeño del modelo en tiempo real y ajustar tanto el preprocesamiento como los algoritmos de modelado conforme se disponga de nuevos datos.

Conclusión Final

El análisis exploratorio, junto con el proceso de preprocesamiento y el modelado, ha permitido desarrollar un modelo predictivo sólido y altamente interpretable. En particular, los modelos lineales, y Ridge Regression en específico, han mostrado un desempeño robusto ($R^2 \approx 0.81$) y ofrecen la ventaja de una fácil interpretación de los coeficientes, lo cual es crucial para la toma de decisiones estratégicas. Sin embargo, la incorporación de datos adicionales y variables externas, junto con la exploración de técnicas de modelado más avanzadas, podría mejorar aún más la precisión del modelo.