

Reporte Final

Modelo de Regresión Lineal con errores estándar robustos para la Predicción del Precio con Descuento

Introducción

Este reporte presenta el desarrollo y evaluación de un **modelo de Regresión Lineal con errores estándar robustos** diseñado para predecir el **precio con descuento** (`discounted_price`) de productos en Amazon. El objetivo principal es optimizar la estrategia de precios mediante la predicción precisa del precio final al consumidor basado en el precio real del producto.

$I = 0.613X + -2.804$. Esto indica que por cada incremento de **1€** en el precio real, el precio con descuento aumenta en **0.613€**.

Descripción de los Datos

El análisis se realizó utilizando un conjunto de datos que incluye las siguientes variables:

- **actual_price**: Precio real del producto en € euros
- **discounted_price**: Precio final del producto después de aplicar el descuento (variable objetivo).

Selección y División de Variables

Se seleccionó una única variable independiente:

- **actual_price**

La variable dependiente es:

- **discounted_price**

El conjunto de datos se dividió en dos subconjuntos:

- **Entrenamiento**: 80% (1,172 observaciones)
- **Prueba**: 20% (293 observaciones)

Evaluación de Multicolinealidad

Se calculó el **Factor de Inflación de la Varianza (VIF)** para asegurar la ausencia de multicolinealidad significativa entre las variables independientes.

Feature	VIF
const	1.253367
actual_price	1.000000

Interpretación: Todos los valores de VIF son inferiores a 5, indicando una **ausencia de multicolinealidad significativa** entre las variables independientes.

Ajuste del Modelo de Regresión Lineal con errores estándar robustos

Se ajustó un modelo de Regresión Lineal con errores estándar robustos utilizando el método de **Mínimos Cuadrados Ordinarios (OLS)** y se aplicaron **errores estándar robustos (HC3)** para manejar posibles heterocedasticidades.

Resumen del Modelo de Regresión Lineal con Errores Estándar Robustos (HC3)

```
OLS Regression Results
=====
=====
Dep. Variable:   discounted_price   R-squared:                0.927
Model:          OLS                 Adj. R-squared:           0.927
Method:         Least Squares       F-statistic:             1191.
Date:           Sat, 02 Nov 2024    Prob (F-statistic):      1.53e-180
Time:           10:29:50            Log-Likelihood:          -5304.8
No. Observations: 1172             AIC:                     1.061e+
04
Df Residuals:   1170              BIC:                     1.062e+
04
Df Model:       1
Covariance Type: HC3
=====
=====
                                coef    std err          t      P>|t|      [0.025
0.975]
-----
const          -2.8037          0.788      -3.557      0.000      -
4.350      -1.257
actual_price    0.6126          0.018     34.505      0.000      0.578
0.647
=====
=====
Omnibus:                387.243    Durbin-
Watson:                  2.046
Prob(Omnibus):          0.000    Jarque-
Bera (JB):              29299.734    Prob(JB):
Skew:                   0.591
Kurtosis:               27.466    Cond. No.
163.
=====
=====
```

Evaluación del Modelo

Métricas de Evaluación en el Conjunto de Prueba

Métrica	Valor
R^2	0.9186
MSE	592.8684
RMSE	24.3489
MAE	9.5791
Durbin-Watson	2.0456

Interpretación:

- **R^2 (0.9186):** El modelo explica aproximadamente el 91.86% de la variabilidad en el precio con descuento, lo que indica un **ajuste muy bueno**.
- **MSE (592.8684) y RMSE (24.3489):** Representan el error cuadrático medio y su raíz, respectivamente, reflejando la precisión de las predicciones del modelo.
- **MAE (9.5791):** Indica que, en promedio, las predicciones del modelo están a aproximadamente 9.58 Euros del valor real.
- **Durbin-Watson (2.0456):** Un valor cercano a 2 sugiere **ausencia de autocorrelación** en los residuos.

Diagnóstico del Modelo

Las pruebas realizadas confirman la presencia de **heterocedasticidad** en los residuos. No obstante, el uso de **errores estándar robustos (HC3)** garantiza la validez de las inferencias sobre los coeficientes del modelo.

Además, la estadística de **Durbin-Watson** cercana a 2 respalda la integridad del modelo al indicar **ausencia de autocorrelación** en los residuos.

Interpretación de los Resultados

El modelo de regresión lineal simple ajustado muestra una **alta capacidad explicativa** con un R^2 de 0.927 en el conjunto de entrenamiento y 0.9186 en el de prueba. El coeficiente de `actual_price` es **0.6126**, indicando que por cada incremento de **1 €** en el precio real, el precio con descuento aumenta en aproximadamente **0.613 € euros**.

Conclusiones sobre el Modelo

El Modelo Regresión Lineal con errores estándar robustos se considera **eficaz** para predecir el precio con descuento de productos en Amazon debido a los siguientes factores:

- **Desempeño Predictivo Superior:** Presenta un alto R^2 tanto en entrenamiento como en prueba, indicando una excelente capacidad explicativa.
- **Coefficiente Realista:** El valor de **0.613** para `actual_price` es coherente y facilita la interpretación práctica del modelo.
- **Ausencia de Multicolinealidad Significativa:** Los valores de VIF son aceptables, garantizando la estabilidad y confiabilidad de los coeficientes.
- **Simplicidad y Facilidad de Implementación:** Un modelo sencillo con una sola variable independiente facilita su implementación y comunicación de resultados.

Modelo Guardado

El modelo final ha sido **guardado** en el archivo `model.pkl` para su uso futuro en predicciones o integraciones en sistemas de producción.

Conclusión Final

El análisis entre `actual_price` y `discounted_price` muestra una **relación fuerte y lineal**. El modelo desarrollado para predecir el precio con descuento (Y) basado en el precio real (X) resulta en la ecuación **$Y = 0.613X + -2.804$** . Esto indica que **por cada incremento de 1 Euro en el precio real, el precio con descuento aumenta en 0.613 Euros**.

Ingresos por Categoría:

- Las categorías con los ingresos más altos son:
- **Electronics** con un total de **€3,138,057.00**.
- **Home&Kitchen** con un total de **€1,042,016.81**.
- **Computers&Accessories** con un total de **€381,720.62**.
- Las subcategorías más lucrativas incluyen **Ordenadores portátiles, Tabletas, Cine en casa, TV y vídeo, Móviles y accesorios, Calefacción, refrigeración y calidad del aire y Cocina y electrodomésticos**.
- La mayoría de los productos están en el rango de **10 a 20€**.

Descuentos por Categoría:

- **Computers&Accessories** ofrece el mayor descuento promedio del **54.02%**.
- **Electronics** ofrece un descuento promedio del **50.83%**.
- La mayoría de los descuentos están en el rango del **40% al 70%**.
- Las subcategorías como **Tecnología portátil** y **Auriculares, auriculares y accesorios** tienen los mayores porcentajes de descuento.

Compromiso del Cliente:

- Las categorías con mayor participación del cliente son **Electronics, Home&Kitchen** y **Computers&Accessories**, representando el **97%** de la variedad de productos y con el mayor número de reseñas.
- Las categorías con mayor satisfacción del cliente son **OfficeProducts** (calificación promedio de 4.31) y **Juguetes y juegos**. La subcategoría **Tabletas** tiene la mayor satisfacción del cliente.
- La mayoría de los productos tienen una calificación de **4 a 4.5 estrellas** y los clientes generalmente permanecen anónimos al dejar reseñas.

Recomendaciones

1.Estrategias de Marketing y Promoción:

- Continuar invirtiendo en **Electronics** y **Home&Kitchen**, las categorías más lucrativas, para mantener y aumentar los ingresos.
- Evaluar la efectividad de las políticas de descuento en **Computers&Accessories** y **Electronics** para optimizar la rentabilidad.

Optimización de Descuentos:

- Analizar si los altos descuentos en **Computers&Accessories** y **Electronics** están atrayendo suficiente volumen de ventas para justificar la reducción de precios.
- Experimentar con diferentes estrategias de descuento para encontrar un equilibrio óptimo entre volumen de ventas y márgenes de beneficio.

Compromiso del Cliente:

- Mantener la calidad de productos y servicios en **OfficeProducts** y **Juguetes y juegos**, donde el compromiso del cliente es alto.
- Investigar y mejorar la calidad de productos en **Car&Motorbike** y otras categorías con calificaciones más bajas para aumentar la satisfacción del cliente.

Análisis de Reseñas:

- Realizar análisis de texto de las reseñas para identificar áreas específicas de mejora y oportunidades de innovación.
- Implementar un sistema de retroalimentación que permita a los clientes expresar sus opiniones de manera más efectiva y responder rápidamente a sus necesidades.

El análisis de los datos de ventas de más de 1000 productos vendidos en Amazon ha revelado que **Electronics** y **Home&Kitchen** son las categorías más rentables, mientras que **Computers&Accessories** ofrece los mayores descuentos. La categoría de **OfficeProducts** tiene la calificación promedio más alta, indicando un fuerte compromiso del cliente. Implementar las recomendaciones mencionadas puede ayudar a maximizar los ingresos, optimizar las políticas de precios y mejorar la satisfacción del cliente en las diversas categorías de productos vendidos en Amazon.