



UNIVERSIDAD SANTO TOMÁS

*Análisis de modelos para la mezcla de
marketing desde la perspectiva de regresión
multiple para una E-Wallet entre los años
2019 y 2022.*

*Tesis de Grado presentado como requisito para la
obtención del título de Profesional en Estadística de
la Universidad Santo Tomás de Colombia.*

Cristian Camilo Pinzón Camacho

Estudiante

CC 1013607813

cristianpinzon@usantotomas.edu.co

Director: Mario José Pacheco López

2022-12

Resumen

En el marketing, la asignación adecuada de los recursos en las distintas formas de darse a conocer a los usuarios finales, se ha convertido en un cálculo en ocasiones difícil de trabajar y entender. Aún más, en los últimos tiempos, con la aparición de nuevas dinámicas a través de canales digitales que aumentan la complejidad del problema.

La estadística ha generado diversos mecanismos para abordar este tipo de problemas analíticos, intentando dar explicación al comportamiento real de la variable de interés. El modelo de mezcla de marketing tiene procedimientos adecuados para exponer las combinaciones que aumentan el rendimiento de los esfuerzos realizados por un Anunciante consignados en diversas variables.

Este trabajo implementa una estrategia que ayuda a definir cuáles son las variables, transformaciones y el modelo, basado en regresión, que permiten encontrar la mejor combinación, dando resultados óptimos del modelo que sean suficientes para predecir la mezcla de marketing más conveniente.

Abstract

In marketing, the appropriate allocation of resources in the different ways of making yourself known to end users has become a difficult calculation to work out and understand. Even more so, in recent times, with the emergence of new dynamics through digital channels that increase the complexity of the problem.

Statistics has generated several mechanisms to address this type of analytical problems, trying to explain the real behavior of the variable of interest. The marketing mix model has adequate procedures to expose the combinations that increase the performance of the efforts made by an advertiser consigned in several variables.

This work attempts a strategy that helps to define which variables, transformations and model, based on regression, allow to find the best combination, giving optimal results of the model that are sufficient to predict the most convenient marketing mix.

1. Introducción

Se denomina mezcla de marketing al conjunto de herramientas y variables utilizadas por los departamentos o áreas de mercadotecnia de las grandes compañías, con el propósito de influir en los clientes finales para obtener los objetivos de ventas trazados con anterioridad.

El término mezcla de marketing fue acuñado por McCarthy en 1960, que reúne cuatro componentes primordiales; producto, precio, distribución y comunicación. Y que se denomina como las 4P's por sus nombres en inglés; *product, price, place and promotion*.

Además, (Peñaloza, 2005) indica que el marketing está en función de las necesidades del cliente y por ende en la consecución de rentabilidad de la empresa. Indica que la necesidad del ser humano es ilimitada y, por tanto, se deben adecuar bienes y servicios para satisfacer dichas demandas.

También, (Peñaloza, 2005) comenta sobre la importancia del marketing en el mundo, convirtiéndose en ideología primordial de todas las corporaciones que persiguen grandes resultados. Con el fin de lograrlos, la toma de decisiones se convierte en una tarea de mayor complejidad, ya que se deben desplegar esfuerzos que consumen gran cantidad de recursos, los cuales se deben seguir y vigilar constantemente, asegurando su óptima distribución.

De otra forma, en un mundo donde proliferan las opciones y las ofertas es de vital importancia saber dónde colocar las inversiones de marketing, para poder llegar eficazmente a los usuarios que buscan un determinado producto. De aquí la importancia que tiene la mezcla de marketing y sus diversas "palancas" como son:

- Las decisiones sobre el producto o servicio.
- Las decisiones sobre la distribución de dicho producto o servicio y la dimensión de cada ofertante.
- Las decisiones sobre la comunicación.
- Las decisiones sobre el precio.
- Decisiones sobre el servicio.

Este último ítem lo incluye (Peñaloza, 2005) entendiéndolo como un conjunto de prestaciones que realiza cada organización desde antes de la venta y hasta después de la misma, generando mayor valor del bien y la complacencia del cliente.

Por otro lado, bajo las nuevas tecnologías y la concepción de mercados innovadores, se crean técnicas que se pueden incluir en la mezcla de marketing, variando, sustituyendo y complementando las 4P's. Siempre y cuando la decisión final afecte de forma positiva a la organización, atrayéndole utilidad, cubriendo las necesidades del mercado y entregando productos y servicios de valor (Thompson, 2005).

Ahora, basados en la relevancia que tiene la mezcla de marketing, nos centramos en la comunicación de los productos o servicios, la trascendencia que tiene el establecer el intercambio del mensaje entre el receptor y el emisor. Y sobre todo lo fundamental que resulta la visibilidad de dicho producto o servicio para cada organización dada una comunicación.

La sociedad ha coexistido con la comunicación, la cual ha mejorado con el tiempo y llegando a ser tan moderna debido a las tecnologías existentes, siendo capaz de llegar a gran cantidad de receptores con un único mensaje. El marketing puede estar en todos los medios, siempre con estrategias desarrolladas para alcanzar al público de la mejor forma posible. Adicionalmente, podemos categorizar los medios de comunicación en dos grupos bajo la perspectiva de los expertos en marketing; medios tradicionales o medios Off y medios digitales o medios On.

2. Problema

De acuerdo a John Wanamaker que comentaba, *"La mitad del dinero que me gasto en publicidad es un desperdicio: el problema es que no sé qué mitad es"*. De donde resulta imprescindible entender ¿qué medios de comunicación?, o mejor, ¿cuáles son los canales que están generando resultados y cuáles no?.

Por consiguiente, es importante generar modelos que permitan identificar que medio y canal mejoran las estimaciones de la inversión del Anunciante, permitiendo una reducción en la incertidumbre, una mejoría en los planes de marketing y la toma de mejores decisiones por parte de los expertos.

3. Pregunta Problema

En este sentido, se busca presentar sí; ¿a través de un modelo, se puede contribuir a que el Anunciante logre distribuir sus recursos de inversión en

medios, de forma más eficiente?

4. Objetivos

4.1. Objetivo General

Explorar distintos tipos de modelos para la mezcla de marketing utilizando los datos de los últimos 3 años suministrados por un Anunciante, que permita evaluar la mejor combinación de inversión en medios.

4.2. Objetivos Específicos

- Evaluar las características de cada una de las variables con métodos descriptivos para identificar aquellas que tienen mayor impacto en la variable respuesta.
- Comparar los distintos modelos utilizando las herramientas estadísticas con sus respectivas pruebas y validaciones.
- Presentar el modelo más eficiente analizando sus resultados de acuerdo al comportamiento de la variable respuesta con sus variables explicativas.

5. Marco Teórico

5.1. El Marketing

Precisando en el término marketing se puede decir que es un proceso que involucra una sociedad y se refleja como un acto administrativo, donde los involucrados sacian sus necesidades al innovar e intercambiar bienes y servicios. Es un conjunto de acciones que se realiza entre el mercado y el cliente. Tipos de acciones como son; compilación de información del mercado, del cliente y competidores, diseño de productos, así como conocimiento de estos y su distribución.

Para que el marketing sea útil debe generar un valor en ambos sentidos, entre el vendedor y el comprador. El vendedor debe conseguir retribución

monetaria por los productos o servicios ofrecidos. Mientras que el comprador, debe tener el sentimiento que su compra ocasiona beneficios importantes para sí y para los cercanos con los que podría compartir aquel provecho. Estos rendimientos bidireccionales se consiguen cuando el marketing es capaz de identificar las necesidades y los deseos existentes en un mercado.

De acuerdo con lo anterior y según (Armstrong y cols., 2013), *“El marketing es el proceso mediante el cual las empresas crean valor para sus clientes y generan fuertes relaciones con ellos para, en reciprocidad, captar el valor de sus clientes.”*

5.2. Importancia del marketing

Su importancia reside en el potencial que existe en cada una de sus técnicas, herramientas y en toda su estrategia, que permite vender y también generar sentido de pertenencia ante lo que significa el negocio. Permite mantener la operación de la organización, cautivando la fidelidad de clientes, utilizando herramientas y estrategias para posicionar en la mente del consumidor dicho producto o marca (Peñaloza, 2005).

Descubrir cuáles son las necesidades o peticiones de los clientes se convierte en una tarea ardua. Para ello, el marketing se apoya en distintas áreas del conocimiento como son; la economía, la psicología, la sociología y la estadística, con el fin de averiguar inicialmente dichos requerimientos de los clientes en un mercado, para luego producir o crear bienes y servicios.

5.3. Funciones del Marketing

Dentro de las funciones fundamentales del marketing están; la investigación del mercado para conocer sus posibles compradores, identificando características demográficas, hábitos de consumo, posibles ingresos y comportamientos. Así como, información de los competidores directos, normativas, reglamentos y leyes de cada mercado y funcionamiento de la economía donde se ofertaran los productos y servicios.

5.4. Proceso de marketing

5.4.1. Marketing Estratégico

Básicamente, se trata del análisis sobre la conveniencia que ofrece el mercado, para ofrecer un nuevo artículo. En otras palabras, a qué comprador se quiere atender, cuál es la capacidad para ello y si este satisface las necesidades, además de tener en cuenta si los clientes están en capacidad de pagar por ello.

5.4.2. Mezcla de Marketing

Consiste en entender en dónde, el cómo y cuándo se debe exhibir el producto o atender un servicio fuertemente apoyado en la publicidad. Sin embargo, sin un plan estratégico de mercadotecnia, las organizaciones no lograrán los objetivos y las inversiones se convertirán en un gasto.

5.4.3. Ejecución del Marketing

Asignar lo correspondiente a los distintos entes que se encargaran de colocar en marcha la estrategia planificada, fijando tiempos, recursos y medios para dirigirlos a su finalidad, utilizando las técnicas y tácticas correspondientes que conlleven al logro de los objetivos.

5.4.4. Control del Marketing

Deben crearse mecanismos que permitan evaluar los resultados del plan establecido y determinar cuán efectivo ha sido. Aquí se realiza la retroalimentación respectiva, evaluando el cumplimiento de los objetivos y tomando correcciones pertinentes de ser necesario.

5.5. El mercado

En un mercado, la función del marketing es fracturar la similitud entre los productos y evidenciar que el artículo ofrecido es el mejor para el cliente.

Según (Armstrong y cols., 2013), señala que: “Desde la perspectiva del marketing, un mercado es el conjunto de todos los compradores reales y potenciales de un producto o servicio. Estos compradores comparten una necesidad o deseo particular que puede ser satisfecho a través de relaciones de intercambio.”

5.5.1. El mercado desde el marketing

Se deben tener en cuenta los análisis realizados previamente en cuanto a la competencia, el producto, las necesidades, la publicidad, calidad y distribución del bien o servicio, teniendo como objetivo la diferenciación y evitando la homogeneidad del mismo. Desde esta perspectiva se han de tomar en cuenta:

- Importancia de la marca.
- Línea de productos.
- Instalaciones.
- Personal.
- Políticas.
- Grado de participación en el mercado total.
- Condicionantes de su misión.
- Conocer la empresa y sus metas.
- Conocer las características del producto a ofertar.
- Satisfacer necesidades de una manera rentable.
- Diferenciar al comprador del consumidor.
- Entender el mercado y a la competencia.

5.6. Modelo de Mezcla de Marketing

En la literatura, por ejemplo en (Pandey, Gupta, y Chhajer, 2021), encontramos que el Modelo de Mezcla de Marketing nos presenta una alternativa desde un enfoque analítico en el cual se utilizan los datos de ventas históricos, cuantificando el impacto de la inversión en marketing contra las ventas en una ventana de tiempo. Este método establece un modelo en donde las ventas

o los volúmenes de ellas, se toman como la variable respuesta. Las variables explicativas resultan de los esfuerzos económicos en el plan de marketing.

Según (Mhitarean-Cuvsinov, 2017), "*MMM es la denominación que se utiliza para cubrir los métodos estadísticos adecuados para el modelado explicativo y predictivo de alguna variable de interés, por ejemplo, las ventas o las cuotas de mercado de una empresa.*"

Además, precisa que es importante el MMM para modelar el comportamiento de la variable de interés e intentar predecirlo. A la vez que se busca descomponer dicha variable en dos factores importantes:

- Volumen base: que son aquellas causadas por estacionalidades y conocimiento de la marca.
- Volumen incremental: que son causados por los movimientos en marketing en un intervalo de tiempo.

Por otra parte, de acuerdo con (Wolfe Sr y Crotts, 2011); "*Existe una desconexión entre la investigación académica y la práctica del marketing. Esta se debe en parte, a la combinación inusual del método de ciencia del marketing y modelos econométricos.*"

Desde el enfoque estadístico, el modelo de mezcla de marketing vincula matemáticamente las actividades de promoción de una determinada marca o producto, a lo largo del tiempo, con una variable respuesta, asociada principalmente a las ventas u otra variable que se requiera medir.

Adicionalmente, Facebook Group presenta un complemento para el software estadístico R llamado Robyn que es un código fuente abierto automatizado para Modelo de Mezcla de Marketing (MMM), y su objetivo es reducir el sesgo humano por medio de regresión Ridge y algoritmos evolutivos, permite la toma de decisiones procesables proporcionando un asignador de presupuesto, curvas de rendimientos decrecientes y permite realizar cambios para una calibración de mayor semejanza con la realidad.

5.7. Modelos de Regresión Multiple, Ridge, Lasso y Elastic Net

En muchos casos, las empresas tienden a realizar inversiones en marketing simultáneamente, provocando posiblemente problemas de colinealidad que se reflejan en los parámetros estimados por los modelos de regresión múltiple,

utilizando mínimos cuadrados ordinarios (OLS, Ordinary least squares). Por lo tanto, es viable abordar regresiones que evitan este tipo de problemas, buscando con ello mejorar las estimaciones que permitan un modelo estable y capaz de predecir de forma confiable (Mhitarean-Cuvsinov, 2017).

5.7.1. Algunos conceptos basicos

Overfitting

El riesgo de sobre-ajustar, o en inglés, de hacer overfitting, sucede cuando al querer construir un modelo, debemos tener en cuenta dos factores: por un lado, queremos que las predicciones que nos dé el modelo sean lo más precisas posible, es decir, que el modelo tenga el menor error posible. Por otro lado, sabemos que las observaciones de nuestro conjunto de datos seguramente tengan errores: o bien errores de medida, o bien o estén influenciadas por algún aspecto que no tenemos en cuenta en el modelo.

Ahora bien, la varianza en un modelo nos dice cómo de variable es el modelo. Cuanta mayor es la varianza, más cambios bruscos, subidas y bajadas veremos en las predicciones del modelo. Si la varianza es muy pequeña, el modelo es muy estable y está infra ajustado. Pero si es muy grande, el modelo es demasiado variable y está sobre ajustado. Buscamos un punto medio entre las dos (Bartlett, Long, Lugosi, y Tsigler, 2020).

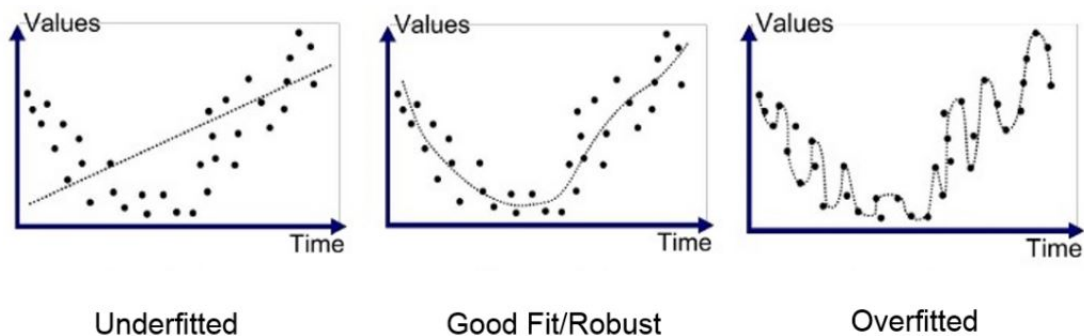


Figura 1: Tomado de: Atomo gris

En los gráficos podemos observar los efectos del sobre ajuste y el sub ajuste en un modelo. Lo ideal siempre será hallar el punto medio.

Sesgo

El sesgo mide la distancia que hay entre el valor esperado de nuestro modelo y el verdadero valor.

$$sesgo(\hat{\beta}) = E(\hat{\beta}) - \beta \quad (1)$$

Básicamente, un modelo de regresión lineal busca que sus estimadores sean insesgados, es decir, el que tiene la menor varianza entre todos los posibles. Sin embargo, que sea el de menor varianza no garantiza que tenga una varianza lo suficientemente pequeña.

Variance bias tradeoff

Si un modelo es demasiado simple y tiene muy pocos parámetros, entonces puede tener un alto sesgo y una baja varianza. Por otro lado, si un modelo tiene una gran cantidad de parámetros, tendrá una varianza alta y un sesgo bajo. Por lo tanto, debemos encontrar el equilibrio adecuado sin sobre ajustar o desajustar los datos (Parast y cols., 2017).

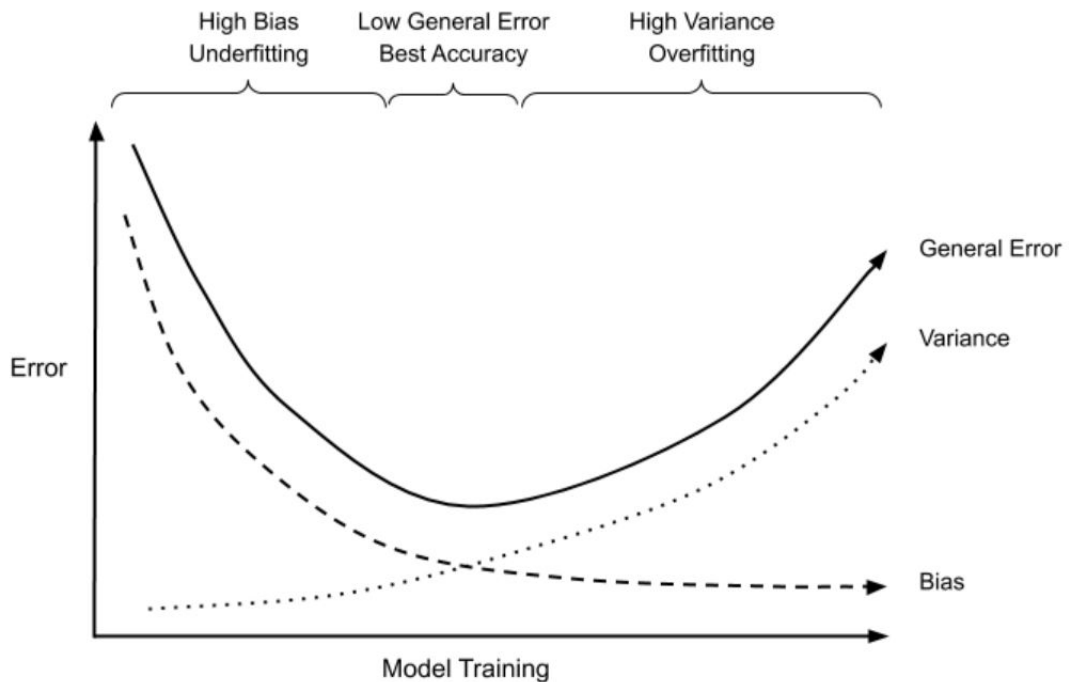


Figura 2: Tomado de: The Science of Machine Learning

5.7.2. Regresión Multiple

Un modelo de regresión lineal para MMM puede tener la siguiente forma:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \dots + \beta_K X_{K,t} + \epsilon_t \quad (2)$$

Tomado de (Mhitarean-Cuvsinov, 2017)

Donde:

- Y_t : Es el valor de la variable dependiente.
- Con $t = 1, \dots, T$ siendo t una observación en el tiempo y el maximo de observaciones T .
- $X_{k,t}$ = Variables independientes
- Con K covariables
- $\beta_0, \beta_1, \dots, \beta_K$, parámetros del modelo.
- y ϵ_t el error.

Siendo $\beta_0, \beta_1, \dots, \beta_K$ desconocidos y el principal objetivo su estimación.

Adicionalmente, se busca que este tipo de modelo cumpla con el teorema de Gauss-Márkov (Carmona, 2005), donde:

- $E(\epsilon_t) = 0$
- $var(\epsilon_t) = \sigma^2 \mathbf{I}_k$ es constante (Homocedasticidad).
- $E(\epsilon_t \epsilon_s) = 0 \quad \forall t \neq s$

5.7.3. Método de Minimos Cuadrados

Es la selección del vector de $\hat{\beta}$ que cumple:

$$\hat{\beta} = \min_a (Y - X\beta)^T (Y - X\beta) \quad (3)$$

Derivando (3) e igualando a cero se tiene la solución de la ecuación y el mínimo de un β cumple con:

$$X^t X \beta = X^T Y \quad (4)$$

Con el estimador de mínimos cuadrados siendo:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (5)$$

En (4) $X^T X$ debe ser no singular para que $\hat{\beta}$ este bien definido (Mejuto Vázquez, s.f.).

Aunque el método es bastante robusto y permite realizar interpretaciones sencillas sobre los β , mostrando buen desempeño predictivo y aproximaciones razonables, por su condición de linealidad no es adecuado al momento de modelar la realidad de los efectos del marketing. Ya que un cambio en $x_{k,t}$ se convierte en un aumento de β_k unidades en y_t implica rendimientos de forma constante. En cambio, los efectos del marketing tienden a presentar una curva de respuesta, exhibiendo un comportamiento no constante o rendimientos decrecientes a escala. Lo cual sucede cuando y_t aumenta con los cambios en las covariables, pero cada unidad adicional de $x_{k,t}$ genera menos en la variable respuesta (Mhitarean-Cuvsinov, 2017).

De acuerdo a la teoría de la ley de los rendimientos marginales decrecientes y los conceptos de marketing, estos se encuentran muy relacionados entre sí. Por un lado, recordemos que el marketing es un acto administrativo donde los involucrados intercambian bienes y servicios. Y por otro, los rendimientos marginales decrecientes, dado el "*ceteris paribus*", representa la disminución progresiva de productos y/o servicios por incluir más factores productivos.

Con lo anterior, se deben plantear alternativas que permitan tratar este tipo de comportamientos e incluirlos en los modelos con el fin de tener mayor certeza y acercamiento a la realidad. Más adelante se estará complementando el tema y mostrando las posibles soluciones.

5.8. Ridge

Este tipo de modelo trata de una regresión con penalizaciones, regulariza el modelo resultante imponiendo una penalización al tamaño de los coeficientes de la relación lineal entre las características predictivas y la variable objetivo.

Son modelos de regresión como el anterior, estimados por mínimos cuadrados, pero que además se incluyen penalizaciones para minimizar la suma de los cuadrados de los residuos, produciendo mejores coeficientes β . Esto se logra al añadir el cuadrado de la norma L2 del vector formado por los coeficientes.

En el caso simple, la fórmula del estimador de mínimos cuadrados:

$$\hat{\beta} = \min_a \sum_{t=1}^K (Y_t - \beta_0 - \sum_{s=1}^P X_{t,s} \beta_s)^2 \quad (6)$$

añadiendo a esta fórmula (5) una penalización en los coeficientes, obteniendo:

$$\hat{\beta}^{\text{ridge}} = \min_a \sum_{t=1}^K (Y_t - \beta_0 - \sum_{s=1}^P X_{t,s} \beta_s)^2 \quad \text{sueto a,} \quad \sum_{s=1}^P \beta_s^2 \leq h \quad \text{con} \quad h \geq 0 \quad (7)$$

restringiendo los valores que toman los coeficientes impidiendo que puedan tomar valores demasiado altos (Mejuto Vázquez, s.f.). Otra forma de ver la restricción es:

$$\hat{\beta}^{\text{ridge}} = \min_a \sum_{t=1}^K (Y_t - \beta_0 - \sum_{s=0}^P X_{t,s} \beta_s)^2 + \lambda \sum_{s=1}^P \beta_s^2 \quad (8)$$

Siendo $\lambda \geq 0$ el parámetro de ajuste, el cual está relacionado con el h de la restricción anterior. Añadiendo el nuevo elemento $\lambda \sum_{s=0}^P \beta_s^2$, coincide con el método de mínimos cuadrados, pero restringiendo los valores que toman los parámetros en función de λ . Si $\lambda = 0$, la penalización es nula y el resultado es equivalente al de un modelo lineal por OLS. Cuando $\lambda > 0$ y aumenta, mayor es la penalización y menor el valor de los predictores, se puede ver fácilmente que cuando $\lambda \rightarrow \infty$, los coeficientes tienden a 0.

La principal ventaja de aplicar ridge frente al ajuste por OLS es la reducción de varianza, el método de ridge es capaz de reducir varianza sin apenas aumentar el bias, consiguiendo así un menor error total.

Cabe resaltar que se debe tener atención cuando se esté asignando el valor de λ , ya que dependiendo de su valor se produce una solución.

Ahora, si las variables explicativas que se están contemplando tienen correlación alta entre sí, los coeficientes no son fáciles de estimar, pero al aplicar la penalización a los coeficientes al cuadrado se impide que ocurra que un coeficiente muy positivo muy grande compense uno negativo. Siendo entonces la regresión ridge una solución a la multicolinealidad.

Se puede notar también que no se aplica esta restricción sobre β_0 , pues la idea es restringir es el efecto estimado que tienen las variables explicativas sobre la variable respuesta.

De forma matricial, la regresión ridge sería:

$$\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T Y \quad (9)$$

Lo que permite corregir el problema de alta varianza, aplicando la matriz identidad con una constante en la diagonal, en caso de que la matriz $X^T X$ sea singular o este cerca de serlo (McDonald, 2009).

5.9. Lasso

Este método combina un modelo de regresión con un procedimiento de contracción de algunos parámetros hacia cero y selección de variables, imponiendo una restricción a los coeficientes similar que la regresión ridge (Ramos Castillo, 2018). El Método Lasso (Least Absolute Shrinkage and Selection Operator) fue introducido por Robert Tibshirani en 1996.

Además, Lasso logra reducir la dimensionalidad de las variables explicativas dado que un predictor con coeficiente de regresión cero no influye en el modelo, consiguiendo excluir los predictores menos relevantes. Al igual que en ridge, el grado de penalización está controlado por el hiperparámetro λ . Cuando $\lambda = 0$, el resultado es equivalente al de un modelo lineal por OLS. A medida que λ aumenta, mayor es la penalización y más predictores quedan excluidos (Mejuto Vázquez, s.f.). La penalización Lasso se conoce también como $L1$.

La forma de restricción es:

$$\hat{\beta}^{\text{Lasso}} = \min_{\beta} \sum_{t=1}^K (Y_t - \beta_0 - \sum_{s=1}^P X_{t,s} \beta_s)^2 \quad \text{sujeto a,} \quad \sum_{s=1}^P |\beta_s| \leq h \quad \text{con} \quad h \geq 0 \quad (10)$$

O de forma Lagrangiana:

$$\hat{\beta}^{\text{Lasso}} = \min_{\beta} \sum_{t=1}^K (Y_t - \beta_0 - \sum_{s=0}^P X_{t,s} \beta_s)^2 + \lambda \sum_{s=1}^P |\beta_s| \quad (11)$$

con $\lambda \geq 0$.

Si quisiéramos verlo de forma matricial, se tiene que (Mejuto Vázquez, s.f.):

$$\hat{\beta}^{\text{Lasso}} = (X^T X)^{-1} X^T Y + \lambda \|\beta\| \quad (12)$$

Se nota claramente la diferencia entre la Regresión Ridge y Lasso en el término $\sum_{s=0}^P \beta_s^2$ y $\sum_{s=0}^P |\beta_s|$. Sin embargo, Lasso tiene la propiedad de anular al menos un β_s siempre que λ sea suficientemente pequeño. Por lo tanto, además de corregir el problema de las variables dependientes, sirve como un método de selección de variables.

5.10. Elastic Net

Elastic net incluye una regularización que combina la penalización $L1$ y $L2$ $\left(\alpha \lambda \|\beta\|_1 + \frac{1}{2} (1 - \alpha) \|\beta\|_2^2\right)$. El grado en que influye cada una de las penalizaciones está controlado por el hiperparámetro α . Su valor está comprendido en el intervalo $[0, 1]$. Cuando $\alpha = 0$, se aplica ridge y cuando $\alpha = 1$ se aplica lasso. La combinación de ambas penalizaciones suele dar lugar a buenos resultados (Zou y Hastie, 2003). Una estrategia frecuentemente utilizada es asignarle casi todo el peso a la penalización $L1$ (α muy próximo a 1) para conseguir seleccionar predictores y un poco a la $L2$ para dar cierta estabilidad en el caso de que algunos predictores estén correlacionados (De Mol, De Vito, y Rosasco, 2009).

5.11. Carryover Effect

Regresando al tema de los rendimientos marginales decrecientes, mencionamos ahora la teoría de la función de producción. Que según (ORTÍZ, s.f.), nos indica que en un sistema económico de mercado, se ocupa de brindar la mayor cantidad de bienes y servicios, utilizando una determinada y finita cantidad de recursos al alcance. Estos insumos otorgados por recursos naturales, humanos y capital son dirigidos con el fin de obtener la máxima ganancia o beneficio.

Además, envuelve la ley de rendimientos marginales decrecientes o principio de productividad marginal física decreciente (Tucker, 2002). Donde expresa que, incrementos adicionales iguales de un insumo variable se van añadiendo a una cantidad constante de un insumo fijo, a partir de cierto punto, los incrementos adicionales del producto total disminuirán (ORTÍZ, s.f.).

Luego, en los Modelos de mezcla de marketing, el principio de productividad marginal se puede modelar a través de ciertas funciones que permiten y dan idea sobre el comportamiento real de las variables explicativa, intentando garantizar que se cumpla:

- Si los gastos son cero, los gastos saturados también.
- Cuanto mayor sea el gasto en insumos, mayor será el gasto en producción saturada.
- Los valores saturados no crecen al infinito, están acotados en algún punto.

También llevando a la pregunta ¿Qué tipo de deterioro debe tener cada variable? O en otras palabras, ¿qué efecto tiene cierta cantidad invertida en una variable en un tiempo determinado después? Esto se conoce como efecto arrastre y aparece cuando algunas de las estrategias de marketing tienen impacto no solo en el período actual, sino también en los períodos futuros (Mhitarean-Cuvsinov, 2017).

En la literatura encontramos que se utilizan funciones como la exponencial de forma tal que:

$$\begin{aligned} y_t &= \beta_0 x_{1,t}^{\beta_1} \epsilon_t \\ \text{con } x_{1,t} &\geq 0 \\ \text{y } 0 &< \beta_1 < 1 \end{aligned} \tag{13}$$

El cual se puede lienalizar de forma:

$$\begin{aligned}
\ln(y_t) &= \ln(\beta_0) + \beta_1 \ln(x_{1,t}) + \ln(\epsilon_t) \\
\text{con } x_{1,t} &\geq 0 \\
\text{y } 0 &< \beta_1 < 1
\end{aligned} \tag{14}$$

Por otra parte, (GitHub, s.f.), comentan sobre las transformaciones de las variables, enlazadas con los principios del marketing como el material publicitario y saturación. Indicando que el Adstock es la técnica que mejor representa el efecto de arrastre de las campañas y la saturación implica que cada unidad adicional de exposición publicitaria aumenta la respuesta, pero a un ritmo decreciente, como se mencionó anteriormente.

Transferencia de Publicidad (Adstock)

Introducido por Simon Broadbent en 1979. explicando la teoría de las acciones publicitarias en el Journal of Market Research Society (Joseph, 2006).

El efecto de arrastre de la publicidad no es inmediato. Específicamente, después de estar expuesto a una campaña de publicidad, no indica que de inmediato se proceda a realizar una conversión o compra de dicha campaña. Además, quiere decir que una vez expuestos a las campañas de publicidad, estas no se olvidan de un día para otro. Aquí es donde cobra vida el remanente publicitario y la teoría del stock publicitario. La forma sencilla de medir de acuerdo a la fórmula de Broadbent:

$$\begin{aligned}
y_t^* &= y_t + \lambda * y_{t-1} \\
t &= 1, \dots, n
\end{aligned} \tag{15}$$

Distribución Weibull

En (GitHub, s.f.) dice que "Ekimetrics & Annalect informó recientemente que la función de supervivencia de Weibull puede adaptarse mejor a la actividad de los medios modernos", y que al establecer los hiperparametros de la transformación dependerá de dos tipos de transformación; la función de distribución y la función de densidad. Teniendo en cuenta que cada hiperparámetro define la forma de la curva.

La función de distribución de una variable aleatoria $X \sim \text{Weibull}(\alpha, \lambda)$ es:

$$\begin{aligned}
F(y) &= 1 - e^{-(\lambda y)^\alpha} \\
\text{para } y &> 0
\end{aligned} \tag{16}$$

Función Hills

Por otra parte, la función Hills permite el modelado para la saturación de los canales de medios con dos parámetros donde uno controla la forma de la curva y el otro el punto de inflexión.

$$y = \frac{z^b}{z^b + K^b} \quad (17)$$

Tomado de (Neubig, Spedding, Kenakin, y Christopoulos, 2003)

5.12. Validación

Los mecanismos para la validación y prueba del modelo comienza con la prueba de los supuestos estadísticos del modelo. La regresión lineal supone que los errores se distribuyen normalmente $\epsilon \sim (0, \sigma^2)$. Para obtener estimaciones puntuales de los coeficientes y realizar inferencias estadísticas como pruebas de significancia o intervalos de confianza.

Para lo cual, se revisan los parámetros de bondad de ajuste del modelo y se establecen las siguientes pruebas:

- Test de normalidad sobre los errores y su respectivo gráfico.
- Durbin Watson
- VIF
- AIC
- RMSE
- R^2

6. Descripción Metodológica

Con el fin de alinear los objetivos del proyecto con la solución del problema, se dispone a realizar un estudio de tipo descriptivo, que nos permitirá medir la tendencia de los datos, encontrando rasgos, patrones y comportamientos que componen las inversiones en medios sobre la E-wallet del Anunciante.

Las variables por utilizar son suministradas por un Anunciante, las cuales están anonimizadas. Se tiene conocimiento que son compiladas por el Anunciante de diversas fuentes de datos, como por ejemplo de su propio sistema de información, datos abiertos de variables económicas y fuentes especializadas.

Estas variables son de tipo numérico y tenemos:

- Total ventas en dólares (variable a modelar).
- Cantidad de unidades vendidas.
- Inversiones en medios de publicidad. (Medios tradicionales y digitales)
- Información macroeconómica. (Inflación, Desempleo, Índice de Producción, Casos COVID, entre otros)
- Cantidad de comentarios sobre el Anunciante en redes sociales.

Esta información está discriminada por semanas, desde la semana 14 del año 2019 hasta la semana 18 del 2022.

Se realizó un análisis descriptivo para intentar hallar aquellos patrones e influencia de las variables que podrían explicar el comportamiento de la E-Wallet en su total de ventas.

6.1. Selección de las variables

Las variables explicativas se probaron todas, sin transformaciones, en modelos iniciales (Lineal, Ridge Lasso y Elastic Net) sin encontrar una combinación óptima que se ajustara a los resultados esperados.

Luego, se transformaron las variables con el fin de darle sentido a la teoría descrita y buscar la mejor combinación. Quedando al final las siguientes:

- Inversión en medios de publicidad Digital
- Inversión en medios tradicionales
- Recordación de la marca
- Comentarios negativos en redes sociales
- Inversiones en publicidad y marca de la competencia
- Casos COVID
- Semanas de Promociones

- Semanas de descuentos y convenios

6.2. Estimación de los modelos

Para los datos de esta E-Wallet se quiere modelar el movimiento de las ventas dado en dólares. Para esto, al incluir las variables anteriormente nombradas se obtuvo los mejores resultados.

Además, se recrean modelos Ridge, Lasso y Elastic Net con estas únicas variables, comparándolos principalmente por su RMSE, arrojando que el modelo Elastic Net es el de mejor estimación.

6.3. Diagnóstico del modelo

Para diagnosticar el modelo final, se emplean las pruebas en cada caso, se evalúa en R_2 y el RMSE se realizan los respectivos gráficos.

Finalmente, se generan las curvas para las inversiones y se da un pequeño análisis respecto a las mismas.

7. Resultados

El comportamiento de la E-Wallet se presenta en el siguiente gráfico:

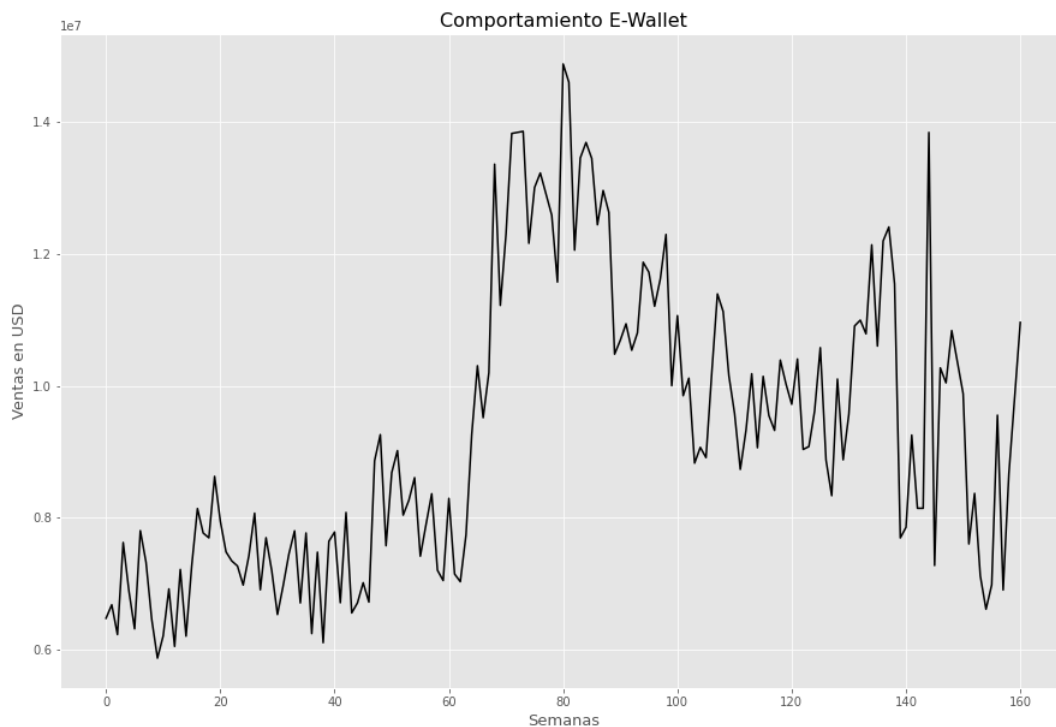


Figura 3: Fuente: Elaboración Propia

La Serie Temporal de movimientos en dólares de la billetera electrónica (E-wallet), consta de 161 datos con periodicidad semanal desde la semana 14 de 2019 (7-04-2019) hasta la semana 18 de 2022 (1-05-2022). El primer análisis de la serie se realiza a través de su gráfico de la serie original, en el cual se extraen las ideas básicas del comportamiento.

De acuerdo a los valores de la serie temporal movimientos de la billetera electrónica en dólares, así como de su gráfica, se observa que existe un cambio significativo a partir de la semana 24 de 2020, periodo que corresponde a la fecha 14-06-2020 u observación 63. Dicho cambio se puede ver afectado por los siguientes factores macroeconómicos:

1. Primer día sin IVA que estaba cerca la fecha de la semana donde empieza el cambio de la serie.
2. 19 de junio de 2022, el pago de primas durante las fechas del 15 de junio a 30 de junio del mismo año.
3. 3 Reapertura de varios sectores de la economía después de cuarentena obligatoria por COVID 19.

4. Expedición del decreto 847 de 2020, el cual regula en el documento, está que los aeropuertos, restaurantes e iglesias del país podrán pedir autorizaciones, a las autoridades locales y nacionales, para implementar planes pilotos de funcionamiento.
5. Jornadas de descuentos por internet.
6. Precios de Avianca para vuelos en septiembre (Apertura de vuelos).

Ahora, presentamos los resultados del modelo lineal múltiple con las variables anteriormente descritas en la sección 6.1 y con sus debidas transformaciones por medio de los efectos de saturación y arrastre.

Nota: Se recuerda que este modelo está hecho con mayor cantidad de las observaciones.

OLS Regression Results							
Dep. Variable:	Venta_USD	R-squared:	0.915				
Model:	OLS	Adj. R-squared:	0.910				
Method:	Least Squares	F-statistic:	204.3				
Date:	Fri, 25 Nov 2022	Prob (F-statistic):	2.88e-77				
Time:	19:15:27	Log-Likelihood:	-2380.0				
No. Observations:	161	AIC:	4778.				
Df Residuals:	152	BIC:	4806.				
Df Model:	8						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	6.846e+06	5.58e+05	12.280	0.000	5.74e+06	7.95e+06	
INVER_DIG_TRANSF	7.194e+06	5.19e+05	13.869	0.000	6.17e+06	8.22e+06	
INVER_TRADIC_TRANSF	2.497e+06	7.4e+05	3.376	0.001	1.04e+06	3.96e+06	
Recordacion_t	6.872e+06	8.18e+05	8.402	0.000	5.26e+06	8.49e+06	
comentarios_negativos_redes_sociales	-3426.3570	541.196	-6.331	0.000	-4495.595	-2357.119	
COMPT_INVER_TRANSF	-4.568e+06	3.05e+05	-14.970	0.000	-5.17e+06	-3.97e+06	
New_cases	8.7077	1.134	7.679	0.000	6.467	10.948	
SEM_PROMOS_DICOT	1.879e+06	1.94e+05	9.665	0.000	1.49e+06	2.26e+06	
DESCUENTOS_Y_CONVENIOS	2.811e+06	2.24e+05	12.528	0.000	2.37e+06	3.25e+06	
Omnibus:	0.979	Durbin-Watson:	2.010				
Prob(Omnibus):	0.613	Jarque-Bera (JB):	0.611				
Skew:	0.091	Prob(JB):	0.737				
Kurtosis:	3.241	Cond. No.	1.15e+06				

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.15e+06. This might indicate that there are strong multicollinearity or other numerical problems.

Figura 4: Fuente: Elaboración Propia

Notamos entonces en esta salida del modelo que:

- Todos sus parametros son significativos siendo menores que el nivel de significancia 0.05
- El DW se encuentra ± 2
- El parametro de bondad de ajuste es muy bueno. $R^2 = 91\%$
- El test de Jarque Bera indica que la hipótesis nula de normalidad no se rechaza.

Tambien, añadimos los resultados del VIF, donde podemos notar que las variables seleccionadas tienen un VIF por debajo de 2, lo que indica que estas variables están poco correlacionadas. Por lo tanto, considerar estas características juntas conducen a un modelo con poca multicolinealidad.

Cuadro 1: Fuente: Elaboración propia

Característica	VIF
Inversión Digital	1.922142
Inversión Med. Trad.	1.396396
Recordación Marca	1.021525
Comentarios Redes Neg.	1.052474
Inversión Competencia	1.176973
Casos Nuevos	1.289771
Sem. Promociones	1.053229
Descuentos y Convenios	1.050100

Ahora se realizan los modelos OLS, Ridge, Lasso y Elastic Net, pero con una partición en la cantidad de variables del 70 % para test.

Los resultados del RMSE se presentan en la siguiente tabla.

Cuadro 2: Fuente: Elaboración propia

MODELO	RMSE
OLS	641773.990
RIDGE	628728.684
LASSO	641773.990
EASTIC NET	611402.132

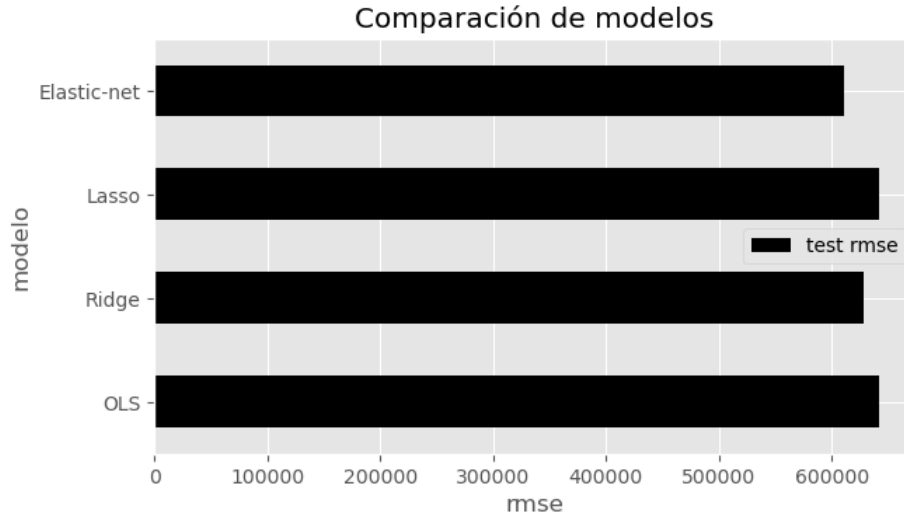


Figura 5: Fuente: Elaboración Propia

Luego, notamos que el modelo por medio del método Elastic Net es el de menor RMSE, y es el que utilizaremos para obtener los resultados de las curvas de saturación de los medios digitales y tradicionales.

Siendo nuestro modelo final:

$$\begin{aligned}
 y_t^{ElasticNet} = & 6857659,82 + 7122182,39(\text{Inversión Digital Transformada}) \\
 & + 2560165,03(\text{Inversión Med. Trad. Transf.}) \\
 & + 7538210,56(\text{Recordación Marca}) - 3818,35(\text{Sentiment Negaativo}) \\
 & - 4546931,04(\text{Inversión Competencia}) + 8,82024(\text{Casos Nuevos}) \\
 & + 1749387,94(\text{Sem. Promociones}) \\
 & + 2723864,95(\text{Descuentos y Convenios})
 \end{aligned}
 \tag{18}$$

Con este modelo realizamos predicciones y comparamos con los datos reales.

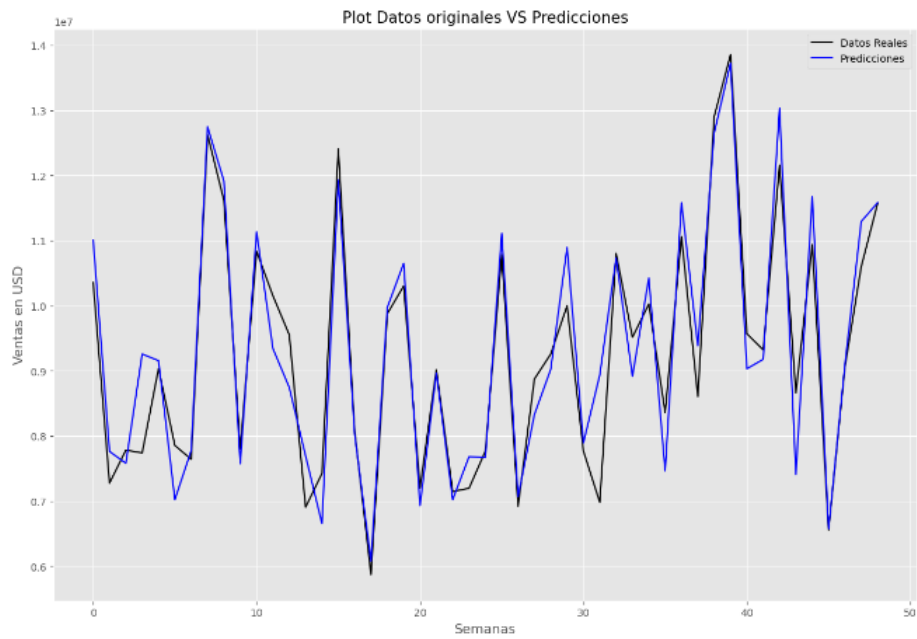


Figura 6: Fuente: Elaboración Propia

Curvas de los medios

Como se planteó en la parte teórica, el comportamiento de los efectos reales en las inversiones de los medios no es creciente de forma lineal. Más bien su comportamiento tiene rendimientos decrecientes marginales. Por lo tanto, la forma de las curvas sería la siguiente:

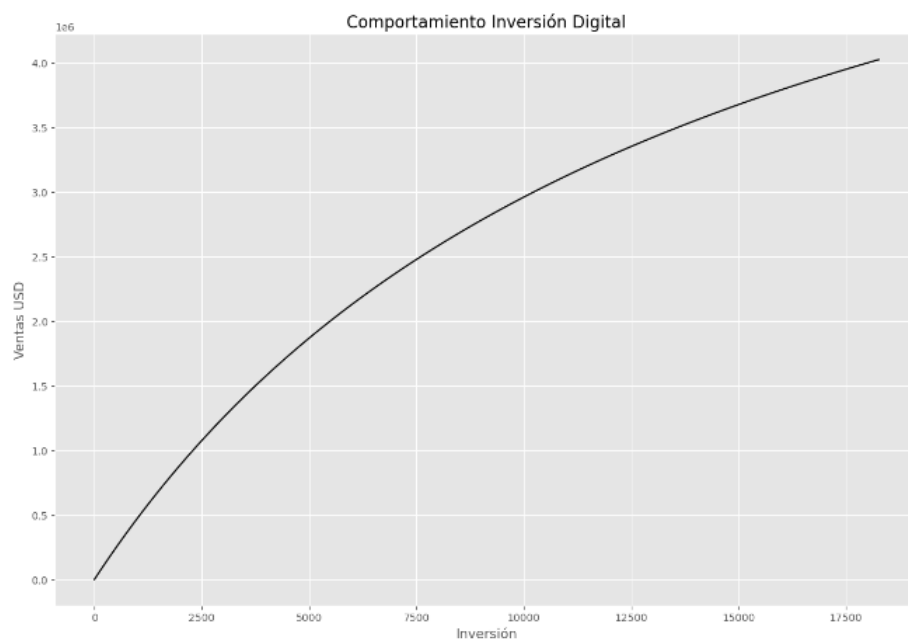


Figura 7: Fuente: Elaboración Propia

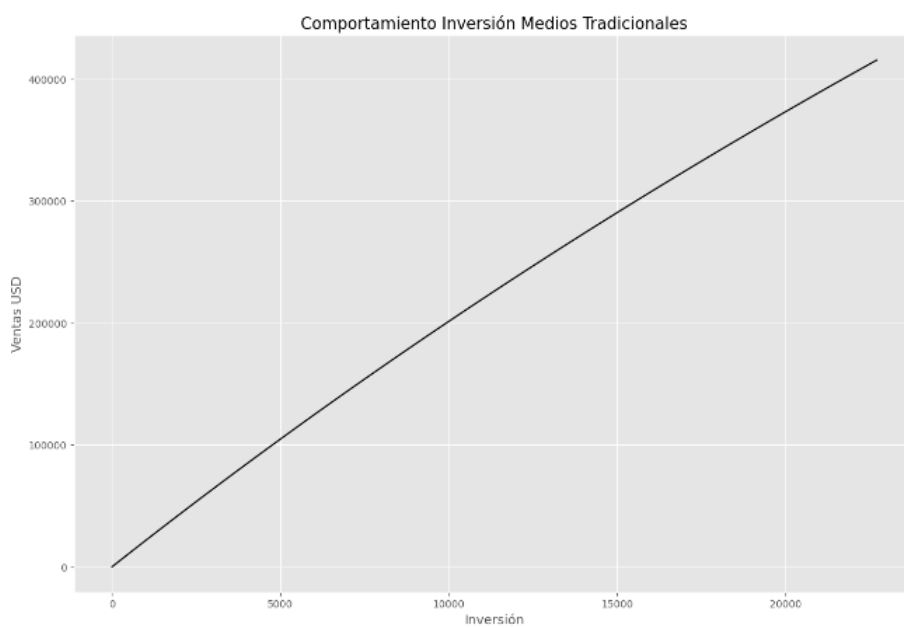


Figura 8: Fuente: Elaboración Propia

Podemos comentar que la inversión en medios tradicionales presenta un comportamiento con rendimientos decrecientes marginales poco marcados. Por

lo cual es viable invertir más en estos medios. Además, es consistente con la interpretación de su coeficiente, siendo este menor al de la inversión digital.

Por otro lado, en los medios digitales vemos que tienen rendimientos marginales más marcados, aunque sugieren que traen mayor cantidad de movimientos a la billetera digital en USD. Esto es consistente con el parámetro del modelo, su interpretación y lo que se esperaba en teoría.

8. Conclusiones

Se logró presentar una aplicación de diferentes enfoques modernos de aprendizaje estadístico en un conjunto de datos suministrado por un Anunciante. El objetivo del mismo fue construir un modelo adecuado para los datos, enfocando el análisis estadístico aplicado a la economía del marketing.

Además, se puede decir que de los modelos presentados en este trabajo, el que presenta mejores resultados es el modelo Elastic Net, ya que su RMSE es el menor. Y con los parámetros calculados por este modelo, se logró realizar los gráficos de curvas de saturación para las variables de inversión de marketing.

Sin embargo, la prueba de los supuestos muestran que las variables explicativas tienen un leve grado de multicolinealidad, sin llegar a ser severa.

Aunque se logra una combinación adecuada de variables y el modelo estima buenos parámetros, es viable trabajar otros modelos que ayuden a mejorar los resultados y los comportamientos de las variables que tienen en teoría un efecto de arrastre.

Además, solamente se trabajó en la búsqueda de la transformación y combinación que diera lugar a un buen modelo. Este tipo de resultados permiten extraer mayor cantidad de información para que las decisiones que tome el Anunciante sean acertadas.

Para finalizar, se recomienda seguir explorando funciones alternativas a las nombradas en este trabajo, que tengan en cuenta los rendimientos marginales decrecientes, las estacionalidades y los efectos de colinealidad que no se abordaron. También, se puede explorar en detalle la librería presentada por Facebook Meta para abordar este tipo de problemas.

Referencias

Referencias

- Armstrong, G., y cols. (2013). Fundamentos de marketing.
- Bartlett, P. L., Long, P. M., Lugosi, G., y Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48), 30063–30070.
- Carmona, F. (2005). Modelos lineales. *Pub. Univ. de Barcelona, Barcelona*.
- De Mol, C., De Vito, E., y Rosasco, L. (2009). Elastic-net regularization in learning theory. *Journal of Complexity*, 25(2), 201–230.
- GitHub. (s.f.). Facebookexperimental/robyn, github. Available at: <https://github.com/facebookexperimental/Robyn> Accessed: November 23, 2022.
- Joseph, J. V. (2006). Understanding advertising adstock transformations. Available at SSRN 924128.
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93–100.
- Mejuto Vázquez, S. (s.f.). Modelos de regresión con penalizaciones.
- Mhitarean-Cuvsinov, E. (2017). *Marketing mix modelling from multiple regression perspective* (Tesis Doctoral no publicada). MA thesis. KTH Royal institute of technology.
- Neubig, R. R., Spedding, M., Kenakin, T., y Christopoulos, A. (2003). International union of pharmacology committee on receptor nomenclature and drug classification. xxxviii. update on terms and symbols in quantitative pharmacology. *Pharmacological reviews*, 55(4), 597–606.
- ORTÍZ, D. E. G. (s.f.). Teoría del consumidor y teoría de la empresa.
- Pandey, S., Gupta, S., y Chhajed, S. (2021). Marketing mix modeling (mmm)-concepts and model interpretation. *Sandeep Pandey, Snigdha Gupta, Shubham Chhajed*.
- Parast, L., McCaffrey, D. F., Burgette, L. F., de la Guardia, F. H., Golinelli, D., Miles, J. N., y Griffin, B. A. (2017). Optimizing variance-bias trade-off in the twang package for estimation of propensity scores. *Health Services and Outcomes Research Methodology*, 17(3), 175–197.
- Peñaloza, M. (2005). El mix de marketing: Una herramienta para servir al cliente. *Actualidad contable faces*, 8(10), 71–81.
- Ramos Castillo, L. (2018). Regresión lasso.
- Thompson, I. (2005). La mezcla de mercadotecnia. I. Thompson, *La mezcla de mercadotecnia*, 2.

- Tucker, I. B. (2002). Fundamentos de economía y microeconomía. *International Thompson, Tercera edición, México*.
- Wolfe Sr, M. J., y Crotts, J. C. (2011). Marketing mix modeling for the tourism industry: A best practices approach. *International Journal of Tourism Sciences*, 11(1), 1–15.
- Zou, H., y Hastie, T. (2003). Regression shrinkage and selection via the elastic net, with applications to microarrays. *JR Stat Soc Ser B*, 67, 301–20.