

**MINIMIZACIÓN DE DEVOLUCIONES EN PORTAFOLIOS DE PUBLICIDAD
CON TÉCNICAS DE APRENDIZAJE AUTOMÁTICO**

DANIEL PATIÑO BARRAZA



**UNIVERSIDAD EAFIT
MAESTRÍA EN CIENCIAS DE LOS DATOS Y ANALÍTICA
MEDELLÍN
2022**

Introducción

Actualmente las organizaciones de retail con estrategias para e-commerce (o ventas virtuales) se ven en necesidad de una toma de decisiones cada vez más analítica para una gran variedad de objetivos. El no estar al día con el uso de metodologías estadísticas o matemáticas para generar o apoyar cuantitativamente las decisiones de los equipos puede generar pérdidas sustanciales en temas como eficiencia, eficacia y en poder de predicción a futuro; lo cual podría verse reflejado como una ventaja a la competencia.

Para este caso específico se explora una empresa orientada a ventas de productos online, que vende los productos de otras empresas, a través de Amazon. Se escoge un subconjunto particular de productos que corresponden a fajas por tener un ratio alto en devoluciones (en comparación con los demás productos) y se plantea la siguiente pregunta: ¿Es posible disminuir las devoluciones de las fajas a través de la modificación de la publicidad hecha para las mismas?, reducir las devoluciones de un producto es importante ya que muchas veces el costo de una devolución del mismo es mayor que el precio de venta del mismo.

Esta idea nace de varios estudios conllevados por la empresa en donde se hipotetiza que las principales razones por las cuales se devuelve una faja son dos y estas son: Tallaje incorrecto para lo cual su solución fue aumentar las fajas dos tallas, es decir, si una faja era talla S, ahora es una talla L y adicional a esto está el uso incorrecto de la faja. La primera logró reducir en gran medida las devoluciones, pero estas siguen siendo un problema preocupante ya que el nivel de devoluciones de las fajas sigue siendo bastante alto

Para tener un contexto más a fondo se presentan los principales usos o categorías de fajas que actualmente se venden:

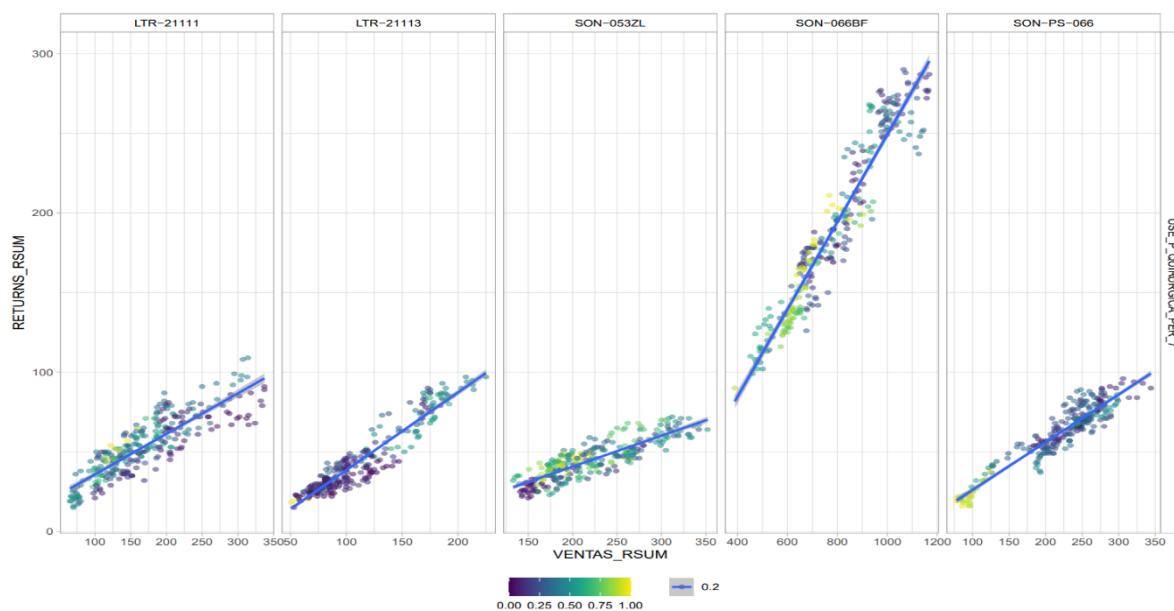
- Post quirúrgicas
- Post Parto
- Uso diario
- Deportivas

Adicionalmente para publicidad existe una categoría extra llamada No_Use_Context, en la cual el tipo de publicidad que se hace es genérico un ejemplo de esto es la publicidad a las palabras: “Fajas Colombianas” y “Fajas”.

¿Porque la publicidad es inherente a este problema? La publicidad en Amazon en su mayoría funciona que los anunciantes escogen que palabras usar para publicitar un producto. Un ejemplo de esto es publicitar la faja 211 con la palabra “faja post quirúrgica” por lo que para cada faja se puede escoger un portafolio de palabras asociada al tipo de uso y esto es controlable por la persona encargada de la publicidad. Para este caso en particular se analizarán las 5 fajas que más venden en Colombia, a continuación, se presenta su composición de publicidad vs su nivel de devoluciones.

	LTR-21111	LTR-21113	SON-053ZL	SON-066BF	SON-PS-066
NO_USE_CONTEXT	35.40%	41.60%	20.90%	46.50%	39.10%
USE_DEPORTIVA_PER	2.00%	2.00%	0.80%	0.30%	1.40%
USE_DIARIO_PER	9.60%	15.00%	40.90%	8.20%	3.80%
USE_P_PARTO_PER	17.40%	5.90%	5.60%	3.20%	5.30%
USE_P_QUIRURGICA_PER	35.60%	35.60%	31.80%	41.80%	50.50%
Devoluciones Promedio	32.00%	38.90%	20.30%	24.00%	26.60%

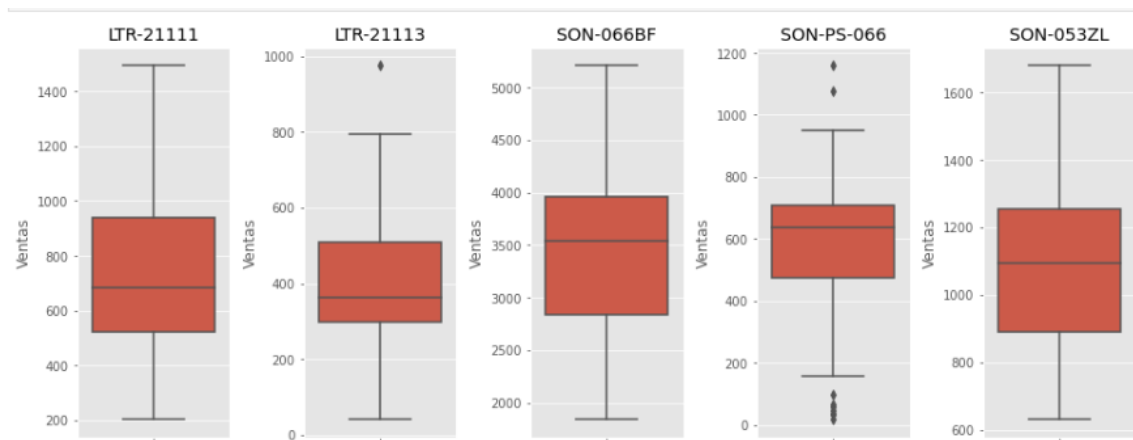
Como ejemplo se revela el impacto que tiene el uso de una categoría en publicidad sobre las devoluciones de cada una de las fajas. Como se puede observar en la siguiente gráfica



en donde se ve la correlación del uso de palabras tipo Post quirúrgicas (medidas en su % del total del portafolio) vs el nivel de devoluciones en las distintas fajas que se analizan en este trabajo, y se ve claramente cómo afecta principalmente a la faja SON-066 (en donde se ve que, a mayor composición de Post quirúrgicas, mayores son sus porcentajes de devoluciones, a diferencia de la LTR 211113 que sucede lo contrario). que es una faja postparto y no postquirúrgica.

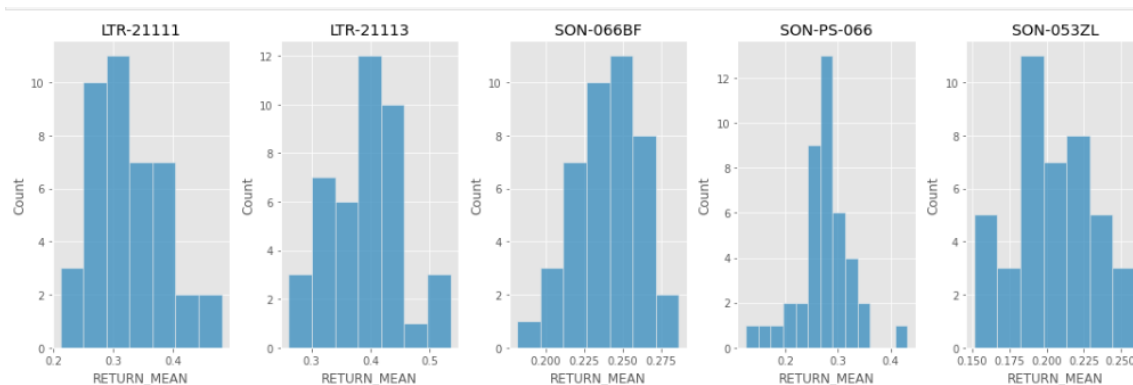
Se venden aproximadamente 7000 fajas por semana, en donde la referencia que más vende es la post quirúrgica (SON-066BF) y se devuelven alrededor del 30% de estas. Se observa que los rangos de ventas por cada faja son bastante variables, por ejemplo las ventas semanales de la faja LTR-21111 varía aproximadamente entre 600 y 1000 fajas mientras la faja SON-053ZL entre 800 y 1200 fajas vendidas cada semana. A continuación se presentan los comportamientos de ventas y devoluciones por cada referencia de faja:

Ventas por fajas



También se revisaron las distribuciones de las devoluciones por cada tipo de faja. Se aprecia que la faja LTR-21113 tiene el porcentaje de devoluciones más grande en promedio con aproximadamente el 40% de fajas devueltas, aunque también varía mucho entre las distintas semanas. Por el contrario el que menos varía es SON-066BF con un promedio de fajas devueltas del 20%. En definitiva el porcentaje de fajas varía aproximadamente entre 0.2% y 0.5%.

Distribución de devoluciones por faja



Ingeniería de características

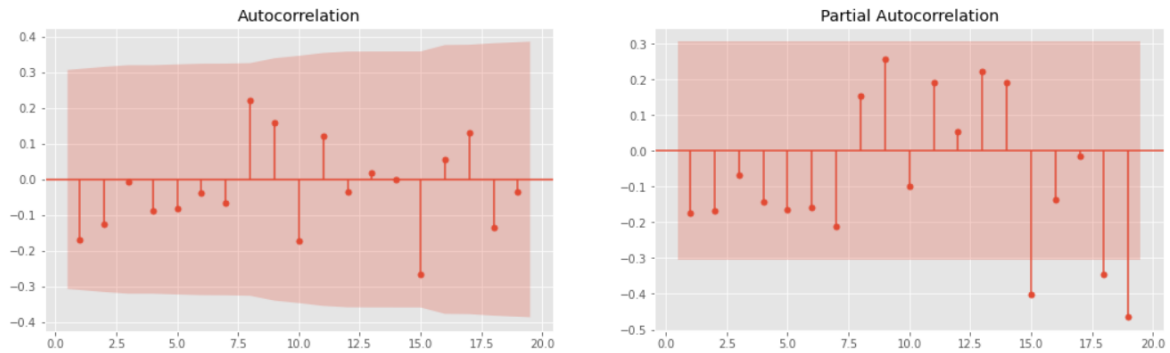
1. Cambio de la frecuencia de los datos

Los datos de las ventas y devoluciones tienen una frecuencia semanal, sin embargo, otras variables como el número de impresiones (número de veces que aparece la publicidad para cada faja) y la cantidad de clicks que hacen los clientes vienen diario, entonces se ajustaron estos datos con frecuencia semanal usando promedios y sumas.

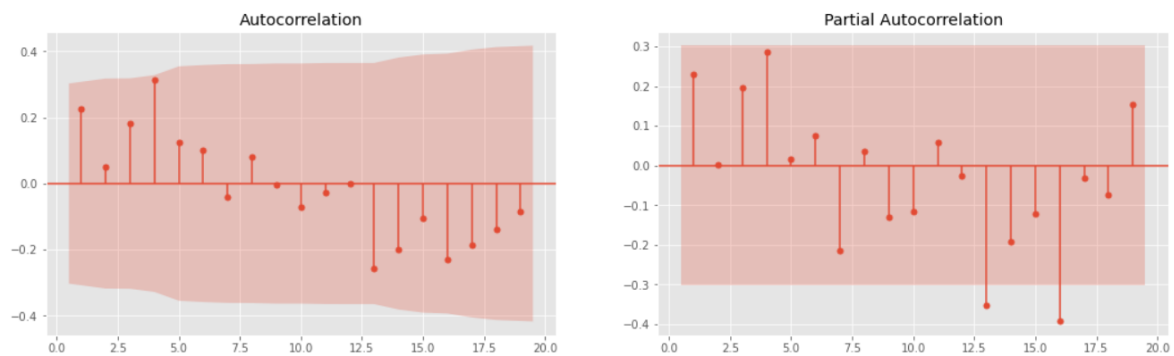
- 2. Variables dummy para representar las fajas:** Se codificaron las distintas fajas usando variables dummy, es decir, 5 variables nuevas con codificación de one hot encoding y se usó este enfoque pues estas fajas no tienen una relación de orden y no se desea que los algoritmos obtengan la falsa impresión de que sí hay una relación de orden.
- 3. Combinación y transformación de atributos.** El número de impresiones equivale al número de veces que aparece la publicidad de cada faja, así que como proceso de ingeniería de características se obtuvo 5 variables nuevas que corresponden con la distribución del portafolio en cada semana, es decir estas variables suman 100% de publicidad en cada semana.
- 4. Selección de variables.** Se tienen variables como el número de impresiones, sin embargo, con las nuevas variables creadas del portafolio de publicidad había alta correlación, entonces eliminamos esta para evitarle problemas a los algoritmos además de que eran explicadas por las otras variables, adicionalmente, se tenía el número de clicks, no obstante, esta no era una variable que se pudiera controlar desde la empresa, así que se decidió excluir y mantener las variables sobre las que se tenía control para hacer la optimización del portafolio puesto que esto va alineado con las necesidades del negocio.

Análisis de series de tiempo en ventas Log Transform followed by Differencing en ventas

Al estar trabajando con una “serie de tiempo” de ventas y devoluciones se analiza que los datos no están autocorrelacionados con sus lags. Se hace una variedad de análisis para encontrar autocorrelación. Entre estos se incluye transformación logarítmica, diferenciación y transformación logarítmica seguida de diferenciación para normalizar los datos y luego evaluarla mediante ACF y PACF:



Análisis de series de tiempo en devoluciones



Cómo se puede evidenciar, tanto ventas como devoluciones no sufren de autocorrelación significativa y esto implica que modelos como ARIMA o SARIMA no se tendrán en cuenta por lo que no hay dependencia secuencial ni de las ventas ni de las devoluciones en el tiempo

Objetivos

- 1) Predecir el nivel de devoluciones para un portafolio de publicidad dado para cada producto
- 2) Predecir el nivel de ventas para un portafolio de publicidad dado para cada producto
- 3) Minimizar el nivel de devoluciones para cada producto usando el modelo 1, y teniendo el modelo 2 para apoyar con información

Enfoque teórico

Para el desarrollo de los objetivos 1 y 2 se necesitaba seleccionar el método de predicción a usar y esto se hizo mediante la prueba de cinco métodos de regresión que son

Regresión de Ridge:

Esta técnica resulta útil como punto de comparación por regular (o penalizar) los coeficientes de variables que pueden no agregar información útil e independiente al modelo. De esta forma es una solución a variables dentro del modelo que de alguna forma tratan de explicar lo mismo. Básicamente ayuda a cerciorarse de que los coeficientes de las

variables que ayudan a predecir se ajusten respectivamente a su importancia dentro del modelo. Adicionalmente al tener el término de penalización agregado a la suma de los residuales al cuadrado, sesga un poco el modelo, lo cual resulta en una varianza reducida y esto estaría reflejado en una predicción más precisa que es indispensable para la fase final de optimización.

Regresión de Lasso:

Similar a la regresión de Ridge, esta técnica se utiliza para tener un punto de comparación importante y puede traer información adicional ya que está (en algunos casos) consigue llevar coeficientes de algunas variables. Esto indicaría que hay variables dentro del modelo que quizás no son necesarias pero el foco principal está en ajustar la compensación de varianza y sesgo para obtener predicciones más precisas que sean útiles para comparar con las otras técnicas.

Random forest:

Esta técnica resulta muy eficaz para predecir ya que el problema en mano es uno supervisado. Esta técnica es buena para alta dimensionalidad y la favorece el utilizar métodos de ensamble como bootstrapping lo cual agrega aleatoriedad durante la fase de muestreo, ayudando a cada variable a converger a un nivel adecuado de importancia dentro del modelo. Adicionalmente combina predicciones de diferentes modelos al mismo tiempo y esto permite que su nivel de predicción sea fuerte; es mejor tener varias predicciones de diferentes modelos que una sola de un modelo.

Extreme Boosting:

Similar a random forest, esta técnica también es un método de ensamble que combina predicciones de diferentes modelos para lograr un modelo de predicción fuerte. La diferencia o el valor agregado que trae es que es secuencial; los árboles de decisión para la predicción se crean de manera aditiva tras correr secuencias de predicciones. Esto puede conllevar a una predicción más fuerte pues cada modelo nuevo trata de mejorar al anterior. Sin embargo, esta técnica tiene una falla importante en términos de explicabilidad o interpretabilidad lo cual sucede por su complejidad y tiene como resultado una respuesta bajo poco contexto que dé seguridad a la hora de tomar decisiones.

Evaluación del modelo

Validación cruzada anidada: La prueba de estos 4 métodos se realizó a través de *validación cruzada anidada*, que lo que hace esta es mostrarnos la bondad de cada algoritmo para generalizar antes distintos conjuntos de datos ya que los prueba bajo muchos re muestreos (lo cual elimina el sesgo de solo probar un conjunto de pruebas). Este enfoque consiste en tener un ciclo interior con validación cruzada para hacer el ajuste de hiper parámetros con las distintas técnicas, pero como este ajuste de hiperparámetros puede sobre ajustarse demasiado al conjunto de datos entregado se usa un ciclo externo de validación

cruzada para evaluar en promedio qué tan bien generalizan las técnicas. Esta metodología reduce la varianza notablemente en la estimación de las métricas de error en comparación con otros enfoques como train-validation-test split y una validación cruzada única, aunque tiene la desventaja de que es más costoso computacionalmente, sin embargo, en nuestro caso en específico es completamente razonable ya que no toma mucho tiempo.

Métricas

Se usaron varias métricas importantes para la evaluación de algoritmos para problemas de regresión como lo es el R^2 que permite comprender la varianza explicada de la variable dependiente usando las variables predictoras, además que es una métrica autocontenida que entre más cerca a 1, implica un mejor resultado del modelo. Se consideran también el RMSE y el MAE como las métricas más importantes en referencia al caso de negocio ya que es indispensable poder cuantificar el posible error en ventas y devoluciones para calcular qué tan costoso le resultaría esto a la empresa.

Es posible observar que el mejor modelo es el Random Forest pues si bien casi todos los modelos tiene un R^2 cuadrado y un RMSE similar en promedio, el Random Forest tiene menor desviación estándar en estas métricas lo que demuestra que es un modelo ligeramente más estable ante cambios en los datos.

- Modelo 1 métricas de la predicción de ventas

model_name	RandomForestRegressor	Lasso	Ridge	XGBRegressor
r2_mean	0.851080	0.845348	0.846941	0.850773
r2_std	0.055356	0.057691	0.051301	0.063332
mse_mean	190979.617733	198296.473634	199258.660886	194162.844975
mse_std	59846.521421	70343.385897	71344.456477	81431.379470
rmse_mean	431.969316	438.778262	439.857401	431.587361
rmse_std	66.197639	75.961243	76.053452	88.854908
mae_mean	304.631762	319.011998	320.337457	308.709901
mae_std	47.067542	55.579125	56.837456	57.625202

Si se incluye la semana como variable predictora es evidente que estos modelos son muchos más poderosos como se aprecia en la siguiente gráfica lo cual puede indicar que el modelo está aprendiendo patrones de las ventas que dependen de la temporada en la que se

está, esto debe ser validado en un futuro cuando se tengan más datos de años anteriores para comprobar este hipótesis

model_name	RandomForestRegressor	Lasso	Ridge	XGBRegressor
r2_mean	0.933297	0.845249	0.847619	0.919631
r2_std	0.030602	0.058130	0.051890	0.032220
mse_mean	88216.064275	197224.640586	197346.416262	108205.959562
mse_std	51178.455451	67898.636160	69165.656678	50521.122641
rmse_mean	286.984457	437.729803	437.800256	320.380577
rmse_std	76.524413	74.948383	75.348208	74.580462
mae_mean	203.685000	321.935810	322.814858	209.398852
mae_std	49.163711	49.604114	51.382016	37.539151

En cuanto a devoluciones, si bien se tiene un menor porcentaje de la varianza explicada lo cuál significa que hay más factores que no se tienen y que influye en la predicción, se puede observar resultados satisfactorios ya que se observa un error de aproximadamente el 5% de las devoluciones con una desviación estándar muy pequeña como se observa con el RMSE del Random Forest que de nuevo termina siendo el modelo más importante. También se ve que a Lasso le va muy mal en comparación a los otros modelos con un error en promedio del 8%

- Modelo 2 métricas de la predicción de devoluciones

model_name	RandomForestRegressor	Lasso	Ridge	XGBRegressor
r2_mean	0.615281	-0.040232	0.594129	0.593874
r2_std	0.149060	0.043830	0.133353	0.163109
mse_mean	0.002315	0.006583	0.002458	0.002478
mse_std	0.000778	0.001318	0.000569	0.001043
rmse_mean	0.047499	0.080664	0.049261	0.048923
rmse_std	0.007697	0.008718	0.005636	0.009208
mae_mean	0.034558	0.065448	0.036602	0.037676
mae_std	0.004863	0.008030	0.004789	0.007992

Optimización de hiperparámetros de los modelos de ventas y devoluciones.

Una vez se escogió Random Forest como técnica para estos dos modelos, se procede a realizar un entrenamiento con todo el conjunto de datos ajustando hiperparámetros a través del uso de una técnica conocida como RandomSearch que permite evaluar diferentes combinaciones de hiperparámetros de los distintos modelos, permitiendo evaluar espacios de búsqueda más extensos y con un tiempo de búsqueda razonable en contraposición con la búsqueda exhaustiva de hiperparámetros.

Los mejores hiperparámetros encontrados para los modelos fueron 300 árboles de decisión, con una profundidad máxima de 39 y cada árbol tiene permitido ser entrenado con un máximo de las 6 variables disponibles si bien esto no quiere decir que todos los árboles se evalúan con todas las variables, precisamente se busca que el conjunto de predictores usados por cada árbol sea distinto con la finalidad de reducir la covarianza entre árboles de decisión lo cual mejora el desempeño del random forest.

- **Minimización de la cantidad de devoluciones**

Nuestro problema de optimización queda definido así

$\text{MinY}(\text{returns}) = F(x_1, x_2, x_3, x_4, x_5)$
 $X_1 = \% \text{ participación de publicidad tipo No_Use_Context}$
 $X_2 = \% \text{ participación de publicidad tipo Uso deportiva}$
 $X_3 = \% \text{ participación de publicidad tipo Uso diario}$
 $X_4 = \% \text{ participación de publicidad tipo Uso post parto}$
 $X_5 = \% \text{ participación de publicidad tipo Uso post quirúrgica}$

Como la función se desconoce se asume que es no lineal, pero se tiene el resultado de esta usando la predicción del modelo dado por el boosting. Se pone las variables por debajo del percentil 75 para garantizar que no haya un cambio muy abrupto

s.a:

$$\begin{aligned}x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\x_1, x_2, x_3, x_4, x_5 &\geq 0 \\x_1, x_2, x_3, x_4, x_5 &\leq \text{percentil } 75 + 1.5 * \text{IQR de cada } x_i\end{aligned}$$

En donde esta programación se repite para cada producto y $F(x_1, x_2, x_3, x_4, x_5)$ es el mismo modelo de Random Forrest pero cambiando la matriz de cuál de todas las variables dummies es 1, dejando las demás en 0, la forma de solucionar este problema de optimización fue mediante el uso del método SLSQP, (Sequential Quadratic Programming

for Nonlinear Least Squares), el cual funciona de la siguiente forma básicamente, dentro de una región de valores x los cuales cumplen las restricciones

$$\mathcal{F} := \{ x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0 \}$$

Se escogen vectores aleatorios de valores x y se plantea una ecuación cuadrática (QP) para cada uno, la cual se soluciona a través de un método Lagrangiano

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \lambda^T h(x) + \mu^T g(x)$$

y da un valor a la función \mathcal{L} , luego esos valores los penaliza o no usando el método de mínimos errores cuadrados (LS), y según la penalización escoge su próximo vector de x hasta llegar al mínimo de la función \mathcal{L} , habiendo resuelto para cada faja como resultado obtenemos los pesos de publicidad que debería tener cada faja y el nivel de devoluciones que tendría ese portafolio de publicidad para cada faja.

- Resultados de la optimización

Situación óptima	LTR-21111	LTR-21113	SON-053ZL	SON-066BF	SON-PS-066
NO_USE_CONTEXT	38.65%	3.97%	48.22%	35.04%	0.00%
USE_DEPORTIVA_PER	8.09%	52.26%	5.97%	0.15%	7.97%
USE_DIARIO_PER	0.00%	0.00%	24.04%	0.00%	21.68%
USE_P_PARTO_PER	0.00%	0.00%	0.34%	0.00%	48.68%
USE_P_QUIRURGICA_PER	53.26%	43.77%	21.42%	64.81%	21.68%
Devoluciones Promedio	32.00%	38.90%	20.30%	24.00%	27.40%
Devoluciones Óptimas	26.19%	32.13%	17.55%	21.66%	20.25%
Disminución de returns	-18.16%	-17.39%	-13.56%	-9.76%	-26.10%

Inicialmente vemos que es posible disminuir en un promedio de 17% las devoluciones de cada uno de nuestros productos, adicionalmente se ve que la forma en la cual el algoritmo logra esto es identificando las categorías las cuales para cada producto detecta que los clientes posiblemente devuelvan el producto más y a esas les quita su participación en el portafolio de cada producto.

Comparacion ventas	LTR-21111	LTR-21113	SON-053ZL	SON-066BF	SON-PS-066
Ventas Promedio	727	404	1084	3477	558

Ventas Minimo devoluciones	1145	442	1101	3130	296
----------------------------	------	-----	------	------	-----

Contrastando el cambio que esto genera con las ventas nos damos cuenta que para algunas fajas esto incluso aumentaría las ventas, pero para otras las disminuiría, pero haciendo la equivalencia total de returns, vs aumento de venta hacer los cambios sugeridos por el algoritmo vemos como hacer los cambios traería mayor utilidad a la empresa.

Bibliografía

Marques, J. P., Cunha, D. C., Harada, L. M. F., Silva, L. N., & Silva, I. D. (2021). A cost-effective trilateration-based radio localization algorithm using machine learning and sequential least-square programming optimization. *Computer Communications*, 177, 1–9. <https://doi.org/10.1016/j.comcom.2021.06.005>

Kraft, D. (1988). A software package for sequential quadratic programming.

Fu, Z., Liu, G., & Guo, L. (2019). Sequential quadratic programming method for nonlinear least squares estimation and its Application. *Mathematical Problems in Engineering*, 2019, 1–8. <https://doi.org/10.1155/2019/3087949>

Antoniou, A., & Lu, W. (2021). *Practical optimization: Algorithms and engineering applications*. Springer.

B. M. Pavlyshenko, "Linear, machine learning and probabilistic approaches for time series analysis," 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), 2016, pp. 377-381, doi: 10.1109/DSMP.2016.7583582.