

Impacto del periodo de rebajas en la previsión de ventas del sector textil

Marina Ramallo Blanco

Máster en Técnicas Estadísticas

Curso 2024-2025

En el presente documento se resume el Trabajo de Fin de Máster titulado “Impacto del periodo de rebajas en la previsión de ventas del sector textil”. No se autoriza la publicación de dicho trabajo en la web del Máster en Técnicas Estadísticas por motivos de confidencialidad de la empresa INDITEX.

La gestión eficiente de las rebajas, también denominadas como **saldo**, ha cobrado una relevancia creciente en la estrategia comercial del sector textil. Tradicionalmente programadas después de temporadas de alta demanda, como la Navidad, las rebajas permiten a INDITEX liberar espacio en los almacenes para nuevas colecciones, mientras que los descuentos sirven como atracción para captar un número elevado de clientes. No obstante, el impacto de las rebajas va mucho más allá de la simple liquidación de inventario: las decisiones que se toman durante los periodos de saldo pueden condicionar de forma directa tanto el volumen de ventas como los márgenes de beneficio. Por ello, una política de descuentos adecuada resulta decisiva para mantener la rentabilidad de la compañía. La correcta planificación del periodo de saldo requiere un análisis detallado que permita anticipar el comportamiento del cliente ante los distintos niveles de descuento.

En este contexto, se estableció como principal objetivo del trabajo lograr proporcionar a INDITEX una herramienta útil para asistir en la toma de decisiones informadas en relación con la elección de precios de saldo, optimizando así el rendimiento comercial durante las rebajas y reforzando su posición estratégica dentro del mercado global de la moda.

Para ello, se partió de una base de datos muy amplia y detallada proporcionada por INDITEX, que recoge información histórica sobre ventas, precios y otros aspectos clave. El tratamiento y preparación de estos datos se realizó utilizando el lenguaje **SQL**, que permitió organizar y filtrar la información de manera eficiente para su posterior análisis mediante el lenguaje **Python**.

Antes de abordar diseño e implementación de modelos de predicción se desarrolló un exhaustivo análisis preliminar, apoyado en herramientas tanto gráficas como estadísticas, y que permitió extraer conclusiones de gran valor para su posterior aplicación a la construcción de los modelos de predicción.

A través del análisis preliminar buscamos garantizar una comprensión progresiva de los datos, desde la panorámica general hasta los matices locales. Realizamos un estudio de la evolución de las ventas en cada **canal de venta** (tienda física y online) a lo largo del tiempo en distintos países o **mercados**. Identificamos similitudes importantes en el patrón de las ventas entre distintos mercados, lo que sugiere la existencia de dinámicas comunes presentes durante el periodo de rebajas en los distintos países del mundo. Estas similitudes justifican la posterior construcción de un único modelo conjunto para todos los mercados, lo que permite aprovechar el volumen total de datos y captar patrones globales.

No obstante, también observamos que el comportamiento de las ventas en cada uno de los canales

es muy distinto: mientras que las ventas online muestran un patrón más explosivo y volátil, las ventas en tienda física siguen una evolución mucho más cíclica y regular. Así, las diferencias estructurales entre canales hacen necesario desarrollar modelos específicos para cada uno de ellos, adaptados a sus particularidades.

A través del análisis preliminar también se abordó un aspecto clave en el contexto del presente trabajo: el estudio del efecto de los descuentos sobre las ventas durante el saldo. Este análisis se apoyó en herramientas para el cálculo de correlaciones y se complementó con técnicas de regresión no paramétrica. Como medidas de la correlación empleamos el **coeficiente de correlación de Pearson** (que detecta asociaciones lineales entre variables) así como el **coeficiente de correlación de Spearman** (que captura relaciones monótonas no necesariamente lineales entre variables). Para visualizar estas relaciones monótonas no lineales de una forma más clara y detallada representamos ajustes mediante **regresión lineal local**, con una selección del parámetro de suavizado basada en técnicas como la validación cruzada y la regla del pulgar.

Pudimos comprobar que, si bien existen variables con una correlación muy elevada con las ventas (como la exposición en tiendas o el stock disponible), los descuentos tienen un efecto adicional claramente positivo muy relevante sobre el incremento de las ventas relativas. Además, se observó que incluso en etapas avanzadas del periodo de rebajas, donde cabría esperar un menor impacto, el descuento continúa teniendo una influencia muy importante sobre el comportamiento del cliente. Gracias a este enfoque, fue posible demostrar empíricamente la relevancia de la estrategia de descuentos y justificar su incorporación como variable clave en los modelos predictivos desarrollados posteriormente.

De este modo, fundamentándonos en los resultados obtenidos, se procedió al diseño e implementación de modelos de predicción basados en el algoritmo **XGBoost** (*eXtreme Gradient Boosting*), una de las implementaciones más potentes y populares del algoritmo de *Gradient Boosting*. Fue desarrollada con el objetivo de ofrecer una herramienta altamente eficiente, escalable y regularizable para problemas de predicción tanto en contextos académicos como industriales. Su éxito se ha consolidado en ciencia de datos, destacando por su elevado rendimiento predictivo y por su capacidad para manejar grandes volúmenes de datos de forma eficiente. La naturaleza de nuestro problema, caracterizado por alta dimensionalidad, esparsidad, interacciones complejas entre predictores, y la presencia de atípicos justifican el uso de la herramienta XGBoost en nuestro contexto.

El desempeño de este tipo de modelos depende en gran medida de una selección óptima de los hiperparámetros. Por ello, se empleó la librería **optuna** para realizar una búsqueda eficiente del conjunto de hiperparámetros que minimizase el error de validación. En concreto, se utilizó como criterio de error a minimizar el **WMAPE** medio en validación cruzada con 5 grupos (**5-fold CV**).

Asimismo, se siguió un detallado proceso de análisis de errores sistemáticos con el objetivo de definir el conjunto final de variables explicativas más relevantes para los modelos. Este proceso, basado en ajustes sucesivos y en la detección de patrones de error persistentes, permitió refinar la selección de predictores.

Una vez definidos los modelos finales para cada canal de venta y seleccionados sus hiperparámetros y variables explicativas, se procedió a su evaluación empírica. Para ello, ajustamos los modelos empleando como datos de entrenamiento los relativos al saldo de verano de 2023 y seleccionamos como muestra de validación los datos del saldo de verano de 2024. Esta fase de evaluación cuantitativa de errores incluyó tanto un análisis de las métricas de error (WMAPE y WMPE) obtenidas en el conjunto de *test* como una comparación frente a un **modelo dummy**. El modelo dummy se basa en replicar las ventas del año anterior a modo de predicciones, de forma que establece una referencia muy simple con la que contrastar cuáles son las mejores aportadas por nuestro modelo.

Los resultados obtenidos fueron, en general, bastante satisfactorios en términos de error, especialmente si se tienen en cuenta los conocimientos previos sobre la complejidad del fenómeno a modelizar.

Al contrastar las métricas obtenidas con la experiencia previa del equipo de la empresa en campañas de saldo anteriores, se observó que los niveles de error alcanzados se sitúan dentro de rangos considerados aceptables para la toma de decisiones operativas.

Además, se evaluó el comportamiento de las predicciones a distintos niveles de agregación (por ejemplo, dando predicciones a nivel de mercado), observándose que, a medida que se consideraban niveles de agregación más amplios, las métricas de error disminuían de forma notable. Así, al agrupar las predicciones se tienden a compensar errores individuales, lo que mejora la estabilidad y precisión del modelo en términos agregados. Esta propiedad abre la puerta a usos adicionales del modelo más allá del objetivo de asistente en la decisión de los precios, como por ejemplo la estimación de la evolución total de las ventas de saldo en un determinado país.

Por último, se consideraron diversos ejemplos de predicción bajo distintos escenarios de precios, mostrando así la aplicabilidad más práctica del modelo. Las simulaciones obtenidas bajo la aplicación de distintos niveles de rebajas permiten una comparación directa del impacto de determinados precios en las ventas esperadas. Esto proporciona una herramienta de apoyo para el equipo de negocio y un buen punto de partida a partir del cual seleccionar el precio que optimice las ventas.