

Impacto del cambio climático en la demanda del sector retail

Toribio Tomeo

Máster en Técnicas Estadísticas
Curso 2024–2025

Por motivos de confidencialidad dictaminados por la empresa colaboradora INDITEX, no es posible la publicación de la memoria del presente trabajo titulado "Impacto del cambio climático en la demanda de sector retail". De este modo, se presenta este documento, que resume su contenido.

Resumen

La crisis climática es uno de los principales problemas a los que se enfrenta nuestra sociedad actual. Nos encontramos ante un momento de cambio y transición económica visible e inequívoco que obliga a todos los agentes partícipes a realizar un cambio estructural en su forma de actuar en la economía, con el fin de minimizar los efectos negativos del cambio climático y buscar la sostenibilidad en el medio y largo plazo.

Una de las industrias más afectadas por la inminente amenaza ecológica es sin duda la industria textil. Las repentinas oscilaciones meteorológicas han provocado cambios palpables en el estilo de vida, la forma de producir y los patrones de consumo en el mundo de la moda, fomentando así un proceso de adaptación y búsqueda de respuestas por parte de tanto empresas como consumidores.

Cambios en la temperatura, acortamiento de ciertas estaciones, alteraciones en la cantidad de horas de sol y desastres meteorológicos son solo algunos de los causantes que llevan a las empresas a tener que recurrir a estrategias complejas de predicción y adaptación ante posibles alteraciones de la demanda del mercado. INDITEX, en su posición como empresa líder del sector, busca impulsar la innovación en el mundo de la moda y liderar la transformación de la industria a través del desarrollo de proyectos orientados específicamente a este fin, así como estableciendo una serie de objetivos sostenibles ambiciosos que ayuden a reducir el impacto de su actividad.

Con esta motivación en mente surgió el presente proyecto, cuyo objetivo central ha sido analizar las ventas y devoluciones de la compañía en relación con diferentes contextos climáticos, con el fin de identificar patrones y posibles pautas explicativas vinculadas al entorno meteorológico. El escrito se articuló en dos partes principales: un análisis exploratorio exhaustivo en el que se analizó el comportamiento de los consumidores respecto a las dos líneas de actividad de la compañía y el desarrollo de modelos predictivos orientados a estimar el volumen de ventas y de unidades devueltas en tiendas físicas.

Previo al análisis, es importante comentar que para la realización del presente estudio la entidad puso a disposición un equipo propio, así como el acceso a múltiples plataformas y repositorios en *cloud*. Desde el departamento de *Analytics* se tuvo acceso a las diferentes bases de datos de la entidad, en

formato SQL, en la plataforma *Snowflake*. Dichos datos fueron aprovechados con posterioridad en la plataforma *Databricks*, donde se desarrolló casi en su totalidad el presente proyecto. El entorno de programación utilizado para el procesado de los datos obtenidos ha sido principalmente *Python*, particularmente, debido a la magnitud de los datos en cuestión, se recurrió a *PySpark*, una API de Python para Apache Spark altamente eficiente en el entorno de datos masivos.

Le experimentación dio comienzo con la elaboración de los conjuntos de datos. Tomando esto en cuenta, se consultaron, cruzaron y transformaron diferentes bases de datos puestas a disposición por la compañía, dando lugar a tres tablas principales: una de ventas, otra de devoluciones y una tercera con información climática. Estas tablas sirvieron como base para la realización del análisis exploratorio, proporcionando una visión inicial de los patrones y relaciones presentes en los datos. Empero, para la fase de predicción se consideró necesario implementar un proceso de *feature engineering*, cuyo objetivo fue generar nuevas variables y enriquecer la información disponible. Este enfoque permitió incorporar al modelo información adicional relevante, con el fin de mejorar la capacidad predictiva de los algoritmos y obtener estimaciones más precisas.

Con las muestras de datos preparadas, se inició el análisis exploratorio, donde, a través de la aplicación de diversas técnicas analíticas y gráficas, se logró evaluar la afectación de la componente climática en la toma de decisiones de los consumidores. Particularmente, y en primer lugar, se logró descifrar el comportamiento las ventas y las devoluciones a nivel global bajo distintas situaciones meteorológicas. Adicionalmente, fuimos capaces de detectar cambios estructurales en la temporalidad de las ventas de tres familias de prendas de temporada: abrigos, camisetas y chaquetas. Seguidamente, se estudiaron dos casos de climatologías anómalas y extremas ocurridas en el territorio nacional: el octubre más cálido de los últimos 30 años en Galicia y la tormenta Filomena. De este análisis se extrajo que, en octubre 2023, las ventas en A Coruña sufrieron cambios bruscos en su estructura, debidos al rápido cambio del comportamiento de los clientes influenciado por esta subida térmica y que, por otro lado, la tormenta Filomena trajo consigo una caída abrupta de las ventas. Este apartado culmina con un breve análisis de las devoluciones en tres localizaciones españolas: A Coruña, Madrid y Sevilla. En este apartado tratamos de descifrar como varía el comportamiento de los clientes bajo diferentes niveles de pluviosidad y temperatura, segmentando el todo por tipología de establecimiento, es decir, diferenciando entre tiendas ubicadas en centros comerciales y tiendas a pie de calle.

En lo tocante al segundo bloque de este trabajo, se desarrollaron y aplicaron varios modelos de *machine learning* enfocados al pronóstico de las unidades vendidas y la cuantía de devoluciones en físico. Los modelos ajustado se basan en el *framework LightGBM*, un algoritmo de *gradient boosting* altamente eficiente, que cuenta con compatibilidad nativa con *Apache Spark*, lo que nos permitió trabajar de forma eficaz con varios conjuntos de datos de elevada dimensionalidad. El diseño contempló el desarrollo de un modelo base que no cuenta con la inclusión de ninguna variable climática y un modelo alternativo, al que convenientemente llamaremos modelo climático, que goza del añadido de dichas variables, para cada una de las líneas de actividad.

Uno de los principales impasses que se abordó en este apartado fue la selección óptima de los hiperparámetros. Dicha problemática incide de manera directa en la eficiencia y la efectividad del algoritmo, motivo por el cual el desarrollo de un modelo de aprendizaje automático adaptado a una problemática específica constituye un proceso complejo y altamente demandante en términos temporales. En este contexto, para la optimización de los modelos asociados al ámbito de ventas, se empleó la librería *Optuna*, un *framework* de optimización automática de hiperparámetros, mientras que, en el caso de los modelos vinculados a las devoluciones, se llevó a cabo una búsqueda manual de estos, guiada por las recomendaciones y directrices proporcionadas por el equipo de devoluciones del departamento de *Analytics* de la compañía. Posteriormente, se ajustaron los modelos empleando las muestras de entrenamiento pertinentes, siguiendo metodologías ligeramente distintas para cada línea de actividad.

Contemplados los hiperparámetros óptimos y obtenidos los modelos finales, pasamos a la fase de evaluación. Para cuantificar la exactitud de las predicciones de cada modelo, se consideraron distintas métricas de error, seleccionándose finalmente el WAPE (Weighted Absolute Percentage Error) como la medida principal. En esta fase logramos exponer y cuantificar los beneficios de la inclusión de las variables meteorológicas, en términos de error. En particular, pudimos apreciar una ligera mejora en los pronósticos de las ventas y devoluciones respecto a los modelos base, resultados que, aunque no hayan sido de amplia magnitud, pueden considerarse apreciables dado el nivel de complejidad de la tarea llevada a cabo. Adicionalmente, se profundizó sobre la distribución del error cometido, desglosando este por tiendas, condiciones climáticas y tipología de establecimiento. De ello, apreciamos que el número neto de tiendas afectadas positivamente por el modelo climático fue notoriamente superior a aquellas que sufrieron un efecto negativo, sobretodo en la rama de ventas. Asimismo, se concluyó que los errores no presentan ningún patrón estacional significativo.

Finalmente, se hizo hincapié en la importancia individual que tuvieron las *features* de índole climatológico en el modelado, donde estas demostraron tener un peso considerable en los pronósticos del modelo. Para este propósito, recurrimos a los valores de Shapley (SHAP), unos instrumentos que nos permitieron comprender la contribución individual de cada variable en las predicciones del modelo. Asimismo, la aplicación de esta metodología contribuyó a mejorar la interpretabilidad del modelo y permitió identificar interacciones relevantes entre las variables, proporcionando información valiosa acerca de posibles umbrales de comportamiento que la compañía podría aprovechar en procesos futuros de toma de decisiones.