

Elaboración de un modelo de pronóstico de ventas de productos de retail – Teamcore

Luciano Fabbri Soto-Aguilar

Concepción, Chile

2020

En el presente trabajo se elabora un modelo de pronóstico de ventas de 413 productos en 7 cadenas comerciales basado en un modelo de boosting de gradiente de efectos mixtos auto-rezagado hasta siete días que considera el efecto idiosincrático de cada producto como efecto aleatorio y los efectos estacionales y promocionales como efectos fijos no lineales. El modelo propuesto fue parametrizado mediante una búsqueda manual de hiperparámetros a través de una validación cruzada del entrenamiento del modelo con datos de abril hasta octubre de 2019 con el pronóstico de las observaciones de noviembre de 2019 para finalmente ser evaluado con las observaciones de diciembre de 2019. El modelo obtuvo un R^2 y un MAPE fuera de muestra ex-post de 0.9 y 60% respectivamente, valores de R^2 fuera de muestra superiores al 0.8 para todos los horizontes de pronóstico menores a 30 días y valores MAPE en torno al 20% para horizontes de pronóstico menores a 18 días.

1. Metodología

Para la elaboración de este proyecto se trabajó sobre datos correspondientes a mediciones temporales de ventas de 413 productos en 7 cadenas comerciales de abril 2019 a diciembre 2019. La información contenida representaba valores de ventas en términos unitarios y de ingresos junto con información promocional general. De esta forma la base de datos correspondía a un total de 1000 series temporales multivariantes en la que cada observación correspondía a una medición temporal de lo descrito anteriormente para un producto en una cadena comercial, lo que representó un total de 251,358 observaciones.

En primera instancia se realizó un análisis exploratorio de los datos para poder determinar y comprender de manera más profunda la naturaleza de las ventas que se modelaron mediante el modelo propuesto. Esto incluyó la identificación de características esenciales de las series de tiempo de los diferentes productos en diferentes cadenas comerciales, la relación cruzada entre el comportamiento de las ventas de un mismo producto en diferentes cadenas y otros aspectos que se consideraron relevantes de acuerdo con los datos disponibles. Una vez conocida la naturaleza de los datos y en función de las limitaciones de diferentes modelos se formuló la propuesta para el modelo de pronóstico de ventas.

Para evaluar el desempeño del modelo, éste fue parametrizado mediante una búsqueda manual de hiperparámetros a través de una validación cruzada del entrenamiento del modelo con datos de abril hasta octubre de 2019 con el pronóstico de las observaciones de noviembre de 2019 para finalmente ser evaluado con las observaciones de diciembre de 2019. Luego se evaluó el desempeño del modelo en datos de entrenamiento y de testeo mediante la utilización de las métricas R^2 y MAPE. Finalmente se hace un análisis exploratorio para determinar de forma más integral el desempeño del modelo y posibles alternativas para mejorar su rendimiento.

2. Resultados

2.1 *Análisis Exploratorio*

Se estimó que el pronóstico de ventas puede ser analizado desde dos perspectivas en términos generales: desde la perspectiva de gestión logística o cadena de suministro, o desde el punto de vista de la estrategia o inteligencia de negocios. Desde la primera perspectiva mencionada se considera que hacer el pronóstico de las unidades a vender es más adecuado, mientras que para la segunda se considera más razonable hacer el pronóstico de las ventas [CLP], esto sobre la suposición de que los valores ‘ventas_clp_dia’ correspondían a la utilidad (dado que la definición de este dato es ambigua). Dado que el efecto promocional puede incrementar de forma drástica las unidades vendidas sin necesariamente incrementar las utilidades (dado a que las promociones afectan, generalmente, la utilidad unitaria de los productos) se considera que el segundo enfoque es más útil para los objetivos de la empresa y por lo tanto se trabaja sobre éste.

La proporción de observaciones en períodos normales y de promoción es bastante balanceada, correspondiendo a un 54% y 46% de las observaciones respectivamente. Para evaluar la diferencia de medias de venta se aplicó la prueba no paramétrica U de Mann-Whitney la que indicó una diferencia estadística significativa. Sin embargo, esto no necesariamente indica que este aumento se debe particularmente al efecto promocional. Este efecto puede ser causado por la coincidencia temporal de las promociones con segmentos estacionales de alta demanda o debido a que promociones recientes estén afectando a productos con efectos significativos de tendencia, por lo que, en estos casos, el aumento de las ventas estaría más asociado a efectos temporales de las series.

En referencia a las cadenas comerciales, las cadenas 1,2 y 3 son aquellas que poseen la mayor cantidad de productos únicos. Por otra parte, desde abril a diciembre de 2019, la mayoría de los productos en las diferentes cadenas estuvieron sujetos a entre 3 y 5 promociones diferentes. De forma similar, cada promoción afectó en promedio a 73 productos distintos, lo que indica que las promociones debiesen ser bastante generales y no diseñadas en función de productos particulares. Es importante recalcar que, dado que no se tenía el detalle de las promociones, no fue posible analizar la elasticidad de la demanda de los productos, por lo que se estuvo restringido a un enfoque más generalizado respecto al impacto de las promociones en las ventas.

Mediante el análisis individual de cada serie, se observa que éstas están altamente correlacionadas en términos cruzados con sus pares de diferentes cadenas comerciales. Junto con esto y a grandes rasgos se observa que la gran mayoría de las series tienden a ser estacionales con una periodicidad mensual. Por otra parte, se observa que el comportamiento entre productos es bastante heterogéneo. En términos del número de observaciones, la gran mayoría de las series se compone de aproximadamente 250 observaciones, mientras la mayoría de los productos posee entre 250 y 750 observaciones. Dado que el número de observaciones por serie y por producto es reducido, la utilización de modelos individuales (por serie/producto) no es recomendada puesto que éstos debiesen tender a sobreajustarse a menos que sean parametrizados de tal manera de configurar modelos de muy baja varianza y alto sesgo.

2.2 *Formulación del Modelo*

En primera instancia se pensó en utilizar modelos estadísticos tradicionales como modelos ARMA, ARIMA, SARIMA, etc. para tener un *benchmark* de desempeño, sin embargo dada la configuración del problema, se considera que la implementación de éstos es infactible; para una aplicación sistematizada de

estos modelos (uno por cada serie) sería necesaria la incorporación de un proceso manual para identificar los modelos apropiados para cada serie además de un proceso iterativo de transformaciones correspondientes de acuerdo al modelo seleccionado y la serie en cuestión, por lo que este enfoque fue descartado.

Considerando lo anterior se propone utilizar un modelo de efectos mixtos basado en boosting de gradiente auto-rezagado hasta siete días que considera el efecto idiosincrático de cada producto como efecto aleatorio y los efectos estacionales y promocionales como efectos fijos no lineales. Este modelo permite el ajuste de un modelo global, sin necesidad de clusterizar o separar las series según su comportamiento o según el producto. Las variables utilizadas son las siguientes:

Tabla 1. Definición de Variables.

Variable	Abreviación	Descripción
$Ventas_{ijt}$	V_{ijt}	Utilidades por venta de productos i en cadena j en período t
Es_promo_{ijt}	P_{ijt}	1, si el producto i en la cadena j está en promoción, 0 en otro caso,
$Día_semana_1_t$	DS_1_t	1, si el período t corresponde al primer día de la semana, 0 en otro caso
...
$Día_semana_7_t$	DS_7_t	1, si el período t corresponde al séptimo día de la semana, 0 en otro caso
$Día_mes_1_t$	DM_1_t	1, si el período t corresponde al primer día del mes, 0 en otro caso
...
$Día_mes_31_t$	DM_31_t	1, si el período t corresponde treinta y uno día del mes, 0 en otro caso

El modelo propuesto es el siguiente:

$$V_{ijt} = f(V_{ijt-1}, \dots, V_{ijt-7}, P_{ijt}, DS_{1_t}, \dots, DS_{7_t}, \dots, DM_{1_t}, \dots, DM_{31_t}, \theta) + Z_i b_i + \epsilon_{ijt}$$

Donde $f(X)$ es la función desconocida estimada utilizando XGBoost, θ es el vector de hiperparámetros correspondiente que parametriza al modelo XGBoost, Z_i es la matriz de covariables de efectos aleatorios, b_i es el vector de efectos aleatorios para el producto i y ϵ_{ijt} el error aleatorio.

2.3 Evaluación del Modelo

El desempeño del modelo fue medido de tres formas; desempeño del pronóstico en los datos de entrenamiento (abril-octubre 2019), en los datos de testeo ex-post (diciembre 2019) y en los datos de testeo ex-ante (diciembre 2019). La diferencia entre las metodologías ex-post y ex-ante, es que la primera hace el pronóstico de los datos de diciembre 2019 con los valores correctos de las ventas rezagadas, mientras que el pronóstico ex-ante corresponde a un pronóstico secuencial. Debido a la forma en la que se formuló el modelo (autorregresivo hasta 7 días), para simular un escenario real de aplicación se debe pronosticar un único día en el futuro y luego utilizar este pronóstico como el valor autorregresivo de primer orden en el segundo día y así sucesivamente. Debido a que los errores se comienzan a propagar en el futuro, se estima que mientras mayor sea el horizonte de pronóstico, menor será el desempeño del modelo. Visto de otra perspectiva, el desempeño ex-post correspondería a un desempeño real en caso de que sólo se realicen pronósticos de un día en el futuro, mientras que el desempeño ex-ante sería una aplicación más probable en donde el modelo se utiliza para hacer el pronóstico de las ventas a lo largo de un horizonte de múltiples días en el futuro.

El modelo posee un rendimiento deficiente en un porcentaje mínimo de observaciones a las que se les denominó *outliers*, lo que se ve reflejado en el aumento sustancial en el MAPE al filtrar el 1% de los errores de mayor magnitud relativa (ver Tabla 2). Son *outliers* en términos del pronóstico, no necesariamente de acuerdo con la distribución de la serie a la que pertenece esa observación. Se recomienda explorar estos puntos de bajo rendimiento en el futuro para así elaborar acciones correctivas; ya sea mediante un ajuste del modelo a datos filtrados de *outliers* o incorporación de variables o transformaciones de variables explicativas que puedan hacerlo robusto a estos escenarios

Tabla 2. Métricas de desempeño.

Métrica	Valor
R^2 entrenamiento	0.90
R^2 testeo, ex-post:	0.90
MAPE entrenamiento	2.37
MAPE entrenamiento con remoción del 1% de los <i>outliers</i>	0.35
MAPE testeo, ex-post	0.60
MAPE testeo, ex-post con remoción del 1% de <i>outliers</i>	0.41

Por otra parte, el modelo obtuvo un R^2 y un MAPE fuera de muestra ex-post de 0.9 y 60% respectivamente, valores de R^2 fuera de muestra superiores al 0.8 para todos los horizontes de pronóstico menores a 30 días y valores MAPE en torno al 20% para horizontes de pronóstico menores a 18 días (ver Figura 1). Dado que los valores de R^2 y de entrenamiento y de testeo ex-post son prácticamente equivalentes, es decir, que el modelo consigue explicar una alta y la misma proporción de la varianza de las ventas de los productos, es posible afirmar que la búsqueda manual de hiperparámetros fue exitosa - configurando así un modelo capaz de generalizar de forma satisfactoria las tendencias de los datos de entrenamiento – y que el modelo se ajustó de forma apropiada evitando así el sobre y sub ajuste.

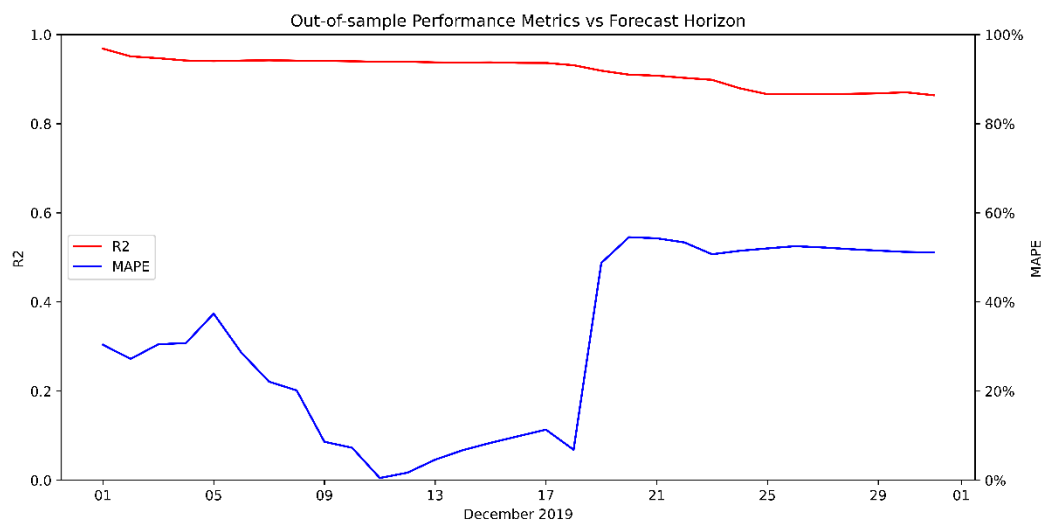


Figura 1. Sensibilidad de las métricas de desempeño fuera de muestra respecto al horizonte de pronóstico.

Por otra parte, se observa que la distribución del error porcentual absoluto es muy similar entre las cadenas comerciales y entre períodos normales y en promoción, por lo que es posible afirmar que el modelo logra generalizar de forma adecuada las tendencias temporales independientemente de las cadenas

y si está en promoción o no. Analizando el pronóstico del modelo por cada serie se identifica que éste pronostica ventas negativas, particularmente cuando las ventas reales fueron iguales a cero. Para trabajos próximos se propone evaluar diferentes técnicas para forzar al modelo a un pronóstico positivo. Para esto se estima que una programación manual del output del modelo podría ser suficiente para escenarios como este.

Para ejemplificar la aplicación real del modelo propuesto, en la Figura 2 se muestra el pronóstico completo del mes de diciembre de 2019. Cabe recalcar que el modelo fue entrenado con datos desde abril a octubre, por lo que este pronóstico es fuera de muestra. La imagen refleja lo expresado en el gráfico de sensibilidad de desempeño en función del horizonte de pronóstico: a partir del día 18, la calidad del ajuste secuencial (ex-ante) sufre un deterioro significativo. El autor considera relevante analizar a futuro tanto la influencia de la propagación de los errores secuenciales como el efecto cíclico de las ventas en el deterioro de la calidad del pronóstico. De todas maneras, una buena alternativa para la utilización de este modelo en el pronóstico real de ventas sería limitar el horizonte de pronóstico a 18 días o menos.

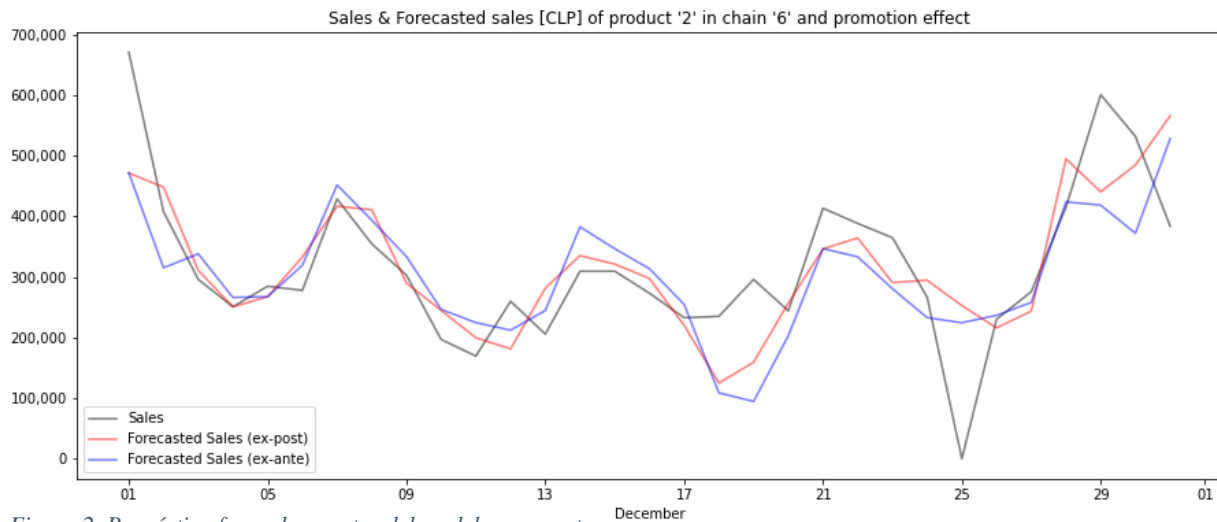


Figura 2. Pronóstico fuera de muestra del modelo propuesto.

3. Conclusiones

Para concluir, el análisis realizado y el modelo propuesto no logran dar respuestas definitivas sobre el impacto de las promociones en las ventas debido a la ausencia de detalles técnicos sobre éstas. Por otra parte, el modelo formulado logró un ajuste satisfactorio generalizando de forma correcta las tendencias temporales de las ventas independientemente de la cadena comercial en cuestión o si el período es promocional o no, sin embargo, mostró un desempeño deficiente para un número reducido de observaciones por lo que se considera necesario analizar las posibles causas para la evaluación de acciones correctivas a futuro. Junto con esto el modelo permite el pronóstico de ventas negativas, principalmente en escenarios en donde la venta real fue igual a cero lo que puede ser corregido sin mayores esfuerzos en siguientes iteraciones. Finalmente, el modelo presenta un desempeño de pronóstico adecuado para horizontes de pronóstico menores a 18 días, lo que lo hace bastante robusto para mercados que presenten una alta frecuencia de abastecimiento de productos, y un desempeño deficiente para horizontes mayores a 18 días por lo que se considera relevante analizar las causas de esta deficiencia para así tener un modelo más robusto que sea aplicable a una mayor gama de mercados comerciales.