Polo del Conocimiento

Pol. Con. (Edición núm. 43) Vol. 5, No 03
Marzo 2020, pp. 92-104
ISSN: 2550 - 682X
DOI: 10.23857/pc.v5i3.1324

Optimización de precios en una empresa de retail utilizando la herramienta rapid miner

Price optimization in a retail company using the rapid miner tool

Otimização de preços em uma empresa de varejo usando a ferramenta de mineração rápida

Diego Heriberto Álvarez-Peralta [I]
dalvarezperalta@hotmail.com
https://orcid.org/0000-0002-9957-5247

Pablo Marcel Recalde-Varela [II]
precalde@uisrael.edu.ec
https://orcid.org/0000-0001-7256-2836

Correspondencia: dalvarezperalta@hotmail.com

Ciencias económicas y empresariales
Artículo de investigación

*__Recibido:__ 30 de enero de 2020 *__Aceptado:__ 28 de febrero de 2020 * __Publicado:__ 10 de marzo de 2020

I. Máster en Dirección de Empresas, Contador Público Autorizado, Ingeniero Comercial con Mención en Administración Financiera, Universidad Tecnológica Israel, Quito, Ecuador.
II. Magíster en Gestión de las Comunicaciones y Tecnologías de la Información, Ingeniero de Sistemas de Computación e Informática, Universidad Tecnológica Israel, Quito, Ecuador.

Resumen

En la presente investigación se empleó la herramienta Rapid Miner Studio que incluye técnicas de Business Analytics y modelos sugeridos por el estado del Arte en el área de Machine Learning para optimizar precios en una línea de productos en la empresa de retail de artículos ópticos más grande en Ecuador. Se incorporaron algunos operadores que son parte del proceso de ETL y limpieza de datos para obtener resultados relevantes que muestran los modelos de mejor desempeño. El estudio es eminentemente cuantitativo por cuanto involucra la aplicación de algoritmos para realizar el análisis predictivo y prescriptivo a los fines de arribar a una solución óptima sugerida en base a Datasets que incluyen atributos definidos por el área de Marketing de la empresa.

Palabras claves: Datasets; preparación; machine learning; modelos.

Abstract

In this research, in order to optimize prices in a product line of the biggest optic retail business in Ecuador the tool Rapid Miner Studio was applied. This tool incorporates Business Analytics and models suggested by the State-of-the-Art. As part of the data preparation, operators suggested by the latest version of the tool were utilized to perform ETL and data cleansing activities. Through this process, it was feasible to obtain valid results based on the best performing models. This research is emphatically quantitative because it involves the application of algorithms to perform both predictive and prescriptive analysis to obtain optimal solutions through the usage of datasets which include attributes defined by the Marketing Area of the business.

Keywords: Datasets; data preparation; machine learning; models.

Resumo

Nesta pesquisa, foi utilizada a ferramenta Rapid Miner Studio, que inclui técnicas e modelos de Business Analytics sugeridos pelo estado da arte na área de Machine Learning para otimizar preços em uma linha de produtos da maior empresa de varejo de artigos ópticos do mundo. Equador Alguns operadores que fazem parte do processo de ETL e limpeza de dados foram incorporados para obter resultados relevantes que mostram os modelos com melhor desempenho. O estudo é eminentemente quantitativo, pois envolve a aplicação de algoritmos para realizar

análises preditivas e prescritivas, a fim de chegar a uma solução ótima sugerida com base em conjuntos de dados que incluem atributos definidos pela área de Marketing da empresa.

Palavras-Chave: Conjuntos de Dados; preparação; aprendizado de máquina; modelos.

Introducción

Como es el caso de muchas empresas pequeñas o de nivel corporativo que pertenecen al negocio de retail, el Grupo OLA a pesar de ser la cadena de ópticas de mayor tamaño en Ecuador, el área de marketing muestra la falencia de no disponer de una política sistemática ni consistente de fijación de precios de su principal producto tangible – los armazones de lentes-.

Al existir en el estado de arte actual varias técnicas de Business Analytics que pueden ayudar en la tarea de fijación óptima de precios, se ha emprendido el proyecto de optimización de precios de manera piloto a fin de contar con un proceso sistemático, soportado en esta rama de la ciencia de datos que permita utilizar una herramienta de vanguardia reconocida por su eficacia a nivel global como Rapid Miner Studio, y que sirva de base de aplicación tanto para la línea bajo estudio, como de referencia para ser aplicadas a otras de la empresa que es objeto de análisis.

El proyecto ejecutado empleó la reportera que es parte del ERP anteriormente implementado por la empresa para obtener información histórica para definir el Datasets necesario para cargarse en la aplicación desde el formato de hojas de cálculo. Al contar con los datos subidos en la aplicación se generaron los procesos de transformación, carga y limpieza de datos para generar a continuación varios modelos de predicción incluidos en el campo de Machine Learning. Sobre aquellos que mostraron el más alto desempeño se procedió a realizar el proceso de optimización obteniendo los resultados sugeridos para su aplicación.

En este sentido, luego de la evaluación de los resultados se obtuvieron conclusiones relevantes que han sido expuestas en su correspondiente sección.

Problema

La fijación de precios de manera intuitiva, no sistemática y que responde solamente a la cobertura de índices de rentabilidad deseado para una línea o a las expectativas de la gerencia, ocasiona que en ciertos casos el precio fijado sea demasiado bajo en relación a la voluntad de pago del cliente y en otros sea demasiado alto en cuanto a los determinantes de la demanda como la percepción

del cliente como el diseño, funcionalidad, durabilidad u otros aspectos hedonísticos como el lujo y la moda. Adicionalmente, si se marcan los precios en base a criterios para alcanzar un mix de ventas y márgenes para cubrir los costos de operación conforme la ubicación de sus locales sugiere que una posible segmentación puede afectar tanto la imagen de las líneas en cuestión como de la empresa.

Objetivos y Justificación

Objetivo General

Optimizar el proceso de fijación de precios para los principales productos de la línea seleccionada mediante la aplicación de una herramienta que incluya las técnicas actuales de Business Analytics (Analítica de Negocios) y Machine Learning (Aprendizaje automático).

Objetivos Específicos

Los objetivos perseguidos en la presente investigación fueron:

- Recolectar la información de fuentes internas como la reportería derivada del sistema ERP mantenido por la empresa para el armado de Datasets, basado en SKUs representativos de una línea de productos.
- Aplicar los operadores necesarios para la preparación de la data y para los procesos de su transformación, cleansing y carga usando la herramienta RapidMiner.
- Validar la pertinencia de los modelos utilizados en cuanto a su nivel de exactitud para sugerir valores óptimos para su solución.

Justificación

Con el uso de técnicas actuales de Business Analytics que aprovechan las capacidades de Machine Learning, la empresa podría gestionar una de las principales variables de Marketing en la que muestra debilidad como el Pricing utilizando los parámetros relevantes que influyan en el logro de sus KPI's (Indicadores claves de desempeño) mediante la aplicación de modelos construidos con datos de entrenamiento para definir y ajustar los mecanismos de predicción que permitan optimizar el ingreso total por ventas de la línea de productos sujeta a estudio.

Métodos

Galkin (2019) sugiere que los algoritmos más avanzados en la arena de Machine Learning considerar un factor relevante para alcanzar un objetivo de maximización. Así, por ejemplo, el

ingreso por ventas. Agrega que en el campo del Deep Learning se utiliza la estimación de la demanda. Éste incluye redes neuronales artificiales que reciben, procesan y transmiten señales a otras neuronas de la red. Cada conexión tiene un peso que determina la fortaleza de dichas señales y el aprendizaje se desarrolla conforme estos pesos son ajustados en base a su desempeño.

Se utilizó el proceso sugerido para la aplicación de problemas de optimización empleando algunos de los tipos de modelos sugeridos por el estado del arte, cuyos resultados fueron evaluados conforme los índices de calidad sugeridos por la herramienta empleada. Sapp (2018) propone las siguientes fases para cumplir con este cometido:

1. Clasificar el problema. - Si corresponde al tipo exploratorio, predictivo, aprendizaje no supervisado que incluye Clustering y K-means o aprendizaje supervisado que se basa en data de entrenamiento e incluye comúnmente redes neuronales, árboles de decisión, redes bayesianas entre otros.
2. Adquirir la data. - Incluye identificar la fuente de data para resolver el problema. Entre ellas están los sistemas ERP, artefactos IoT (Internet de las cosas) o data del mainframe. La data puede ser estructurada como la proveniente de registros No SQL, o no estructurada.
3. Procesar la data. - Incluye la transformación, y cleansing de la data, así como la selección de los Datasets para entrenamiento en el aprendizaje supervisado.
4. Modelar el problema. - Determinar los algoritmos a ser utilizados para entrenamiento o clustering.
5. Validar y ejecutar. - Es probable que comprenda varios ciclos de ejecución de las rutinas y refinar los resultados.
6. Desplegar. - Para crear valor en términos de negocio, éste se manifiesta como data para toma de decisiones, alimentar aplicaciones o ser guardados en un repositorio para análisis a futuro. También puede adoptar el formato de nuevos modelos o rutinas que complementen los sistemas actuales como modelos predictivos. Sea cual sea el formato de salida, en esta fase se determina cómo y cuándo será lanzada para consumo o toma de decisiones.

Todos los modelos empleados en el presente proyecto corresponden al tipo supervisado.

La unidad de estudio correspondió a la línea de armazones de lentes de la marca Guess en la empresa seleccionada. Esta selección fue intencional y sugerida por la dirección de la empresa por cuanto -según su criterio- muestra en sus clientes actuales y potenciales propensión a aceptar un plus en los precios gracias a aspectos hedonistas como el status de la marca y la oportunidad de diferenciar precios gracias a la vinculación emocional que puede derivarse de contar con una marca bien posicionada.

La data seleccionada para el análisis se extrajo principalmente de los módulos del ERP y de la reportería previamente desarrollada por la empresa que se encuentra en su sistema integrado (SAP).

La unidad mínima de estudio es el producto a nivel de SKU (Código de artículo usado por minoristas para identificar y rastrear sus inventarios). La línea de productos sujeta a análisis tiene una cartera de 24 artículos a nivel de SKU conforme a los reportes de cierre de ventas.

Los Datasets utilizados incluyeron información por un período de dos años. A los fines de realizar las cargas iniciales de datos para fines de entrenamiento de los modelos se emplearon originalmente series de tiempo mensuales de 18 meses y posteriormente 24 meses. El primer intento fue fallido dado que se requieren al menos 100 registros y por tanto se utilizaron series de tiempo de períodos semanales dando un total de 104 semanas registradas en cada Datasets empleado.

Dado que los productos de la línea no muestran alta frecuencia en su transaccionalidad bajo análisis éstos no muestran valores de venta en todas las semanas del período. Por ello, la estrategia de reemplazo de los “missing values” fue de total relevancia para obtener modelos que muestren índices de desempeño aceptables minimizando el sesgo y posibilidad de error en la predicción y prescripción.

Para la aplicación del proyecto se utilizaron los lineamientos sugeridos por la metodología de Pmbok en cuanto a gestión de recursos e interesados. En el primer caso, mediante la elaboración de una matriz de responsabilidades Raci se definieron los aportes de los principales miembros involucrados de la empresa como el gerente de marketing, la jefa de sistemas y el gerente general. En el segundo se gestionó el involucramiento de los interesados mediante reuniones de seguimiento quincenales.

Análisis de Resultados

Se incluyó una serie de tiempo de 2 años empleado empleando datos semanales debido a que el proceso de modelado exigió la incorporación de al menos 100 registros para su ejecución.

En el proceso de transformación se utilizó la opción de renombrar atributos para evitar confusión entre descripciones similares como fue el caso de unidades vendidas.

A pesar de existir otras opciones en el proceso de transformación como el uso de filtros, rangos o muestreos de datos, estos no fueron incorporados a los operadores de procesos por la limitada disponibilidad de registros que se explica posteriormente.

Similarmente la opción de ordenamiento que es parte del proceso de transformación tampoco fue aplicada debido a que los reportes extraídos del ERP mostraron la serie de tiempo en forma cronológica y por tanto no hubo necesidad de aplicarlo.

En el proceso de cleansing se eliminaron inicialmente dos atributos como el código de SKU y la marca por su elevada estabilidad y su poca contribución en calidad de predictores.

A pesar de utilizar la opción de eliminación de atributos de alta correlación utilizando un umbral del 80%, no se eliminaron atributos adicionales por no existir descriptores que superen dicho límite.

El total de registros completos utilizados en los Datasets contienen información original que oscila entre el 50% y 70% del total de 104 registros utilizados para fines de modelación.

Para completar los Datasets se utilizaron las siguientes estrategias que se muestran a continuación:

Estrategia de tratamiento de los "Missing values"

Clase	Tipo de Dato	Atributo Original	Reemplazado por:
Categoría	Polinomial	Zona	Más frecuente
Número	Entero	Unidades Vendidas	Valor mínimo
Número	Real	Ventas	Promedio
Número	Real	Precios	Promedio
Número	Real	Utilidad	Promedio
Número	Real	% Margen	Promedio
Categoría	Polinomial	Rentabilidad	Más frecuente

Pol. Con. (Edición núm. 43) Vol. 5, No 03, marzo 2020, pp. 92-104, ISSN: 2550 - 682X

A continuación, se agregaron dos atributos que fueron calculados en base a atributos incluidos en el Datasets original mediante la opción "*Generate*" como sigue:

1. Cálculo de rotación del inventario en veces por año.
2. Inclusión de un nuevo atributo Rentabilidad con valores numéricos (1 o 0) a los fines de incluir variables Dummy así como obtener estadísticos de utilidad.

Posterior a utilizar la vista de resúmenes estadísticos de datos, se generó en la vista de diseño el encadenamiento de procesos a fin de visualizar los operadores utilizados en el proceso de preparación de la data previo a la generación de modelos.

Como paso previo a confirmar los Datasets definitivos, la herramienta sugirió eliminar los atributos de más baja calidad tomando en cuenta principalmente los criterios de correlación y estabilidad.

En los Datasets empleados se generaron típicamente seis tipos de modelos sugeridos, los cuales fueron:

- Lineal generalizado
- Deep Learning
- Árboles de decisión
- Random Forest
- Gradient Boosted Trees
- Support Vector Machine

Goodfellow et al. (2016) señalan respecto al modelo de Deep Learning que éste se basa en encontrar los parámetros que reducen una función de costo o pérdida e incluye una medida de desempeño que es evaluada sobre toda la data de entrenamiento. Agrega que la pérdida esperada en la Datasets de entrenamiento se minimiza al reemplazar la distribución real con la empírica. Sin embargo, advierte que este abordaje puede conducir a un sobre ajuste, y por ello se recurre a descomponer la función objetivo en una suma de ejemplos de entrenamiento.

En cada caso, se seleccionaron los dos tipos de modelo que mostraron el mejor desempeño considerando los criterios del menor error relativo y más alta correlación.

Sobre ellos se ejecutó el proceso de prescripción utilizando el precio como atributo predictor y el kpi de Ventas netas como objetivo a maximizar a fin de preparar dos escenarios. En el primero,

se incluyeron las restricciones globales al actuar dentro del límite de dos desviaciones estándar y los máximos y mínimos registrados en los Datasets base, mientras que en el segundo se ejecutaron sin ninguna restricción.

La herramienta genera los modelos predictivos en cuestión de pocos segundos o microsegundos, siendo el tiempo uno de los parámetros en que miden su desempeño.

Al respecto se ha demostrado no solo que se logra arribar a soluciones óptimas en tan poco tiempo, sino también que se supera los abordajes basados en la clásica heurística tanto en muestras experimentales que contienen datos reales y sintéticos (Bertsimas & Dunn, 2019).

A continuación, se muestran los resultados de los escenarios de precios sugeridos de tres de los sku más representativos de la línea bajo análisis.

Tabla 1: Resultados de los modelos de mejor desempeño Sku 664689793068

Tipo de Modelo	**Con restricciones**			**Sin restricciones**	
	Precio	*Kpi(Ventas)*		*Precio*	*Kpi(Ventas)*
Random Forest	123.11	210.99		125.39	222.51
Deep Learning	123.20	311.83		137.32	309.89

Nota: *El Kpi a lograr según la predicción del modelo corresponde al período de una semana y los valores se expresan en US$.*

Tabla 2: Resultados de los modelos de mejor desempeño Sku 146GU188900953

Tipo de Modelo	**Con restricciones**			**Sin restricciones**	
	Precio	*Kpi(Ventas)*		*Precio*	*Kpi(Ventas)*
Decision Tree	71.71	94.52		81.72	128.05
Random Forest	135.10	136.41		157.65	137.39

Nota: *El Kpi a lograr según la predicción del modelo corresponde al período de una semana y los valores se expresan en US$.*

Tabla 3: Resultados de los modelos de mejor desempeño Sku 146GU745832B58

Tipo de Modelo	Con restricciones			Sin restricciones	
	Precio	*Kpi(Ventas)*		*Precio*	*Kpi(Ventas)*
Support Vector Machine	101.00	99.18		548.72	434.59
Gradient Boosted trees	104.33	105.08		110.71	103.51

Nota: El Kpi a lograr según la predicción del modelo corresponde al período de una semana y los valores se expresan en US$.

Discusión de los Resultados

Tanto al inicio como al final del proceso de cleansing la herramienta brindó la posibilidad de eliminar atributos de baja calidad o de alta correlación para eliminar aquellos que contribuyan a crear multicolinealidad.

Debido a que los índices de rotación de los SKU bajo análisis no se mostraron como atributos(Descriptores) de mayor relevancia para alcanzar el kpi objetivo, se infiere que aún, cuando algún objetivo de rotación de inventario no sea plenamente alcanzado, los esfuerzos derivados de acciones conjuntas de Marketing por sostener el precio óptimo deberían primar salvo que exista un efecto significativo negativo en el flujo de efectivo de la empresa, situación que al momento del análisis no experimentó la empresa.

En los casos observados, se mostró el precio como uno de los factores predictivos de mayor preponderancia siendo además el de mayor peso en los modelos que alcanzaron el mejor desempeño.

Conclusiones

- A través de la aplicación de Rapid Miner Studio se han obtenido valores sugeridos de optimización de precios para sus productos más representativos bajo los escenarios de inclusión y exclusión de restricciones globales.
- En series de tiempo, la herramienta requiere incluir al menos 100 registros como paso previo a la generación de modelos de predicción y prescripción y por tanto se utilizaron 104 registros en cada Datasets que corresponde a un período de dos años.
- La estrategia de reemplazo de Missing valúes es de suma importancia para garantizar la calidad de la data de entrada y del ajuste de los modelos predictivos.

- La opción de eliminación de atributos con alta correlación permite eliminar el fenómeno de multicolinealidad.
- Los reportes de factores de predicción y de pesos en los modelos han mostrado la relevancia del precio como el atributo controlable (descriptor) más destacado entre los utilizados en el presente proyecto especialmente en los SKU que registran mayor transaccionalidad.
- El uso de restricciones permite contrastar los resultados para fines de toma de decisiones sobre la aplicación de resultados. Su aplicación dependerá principalmente del nivel de aversión al riesgo de los tomadores de decisiones.

Recomendaciones

Extender el proyecto hacia líneas de producto que muestren mayor transaccionalidad en su historial a fin de contar con el suficiente número de registros para aplicar el análisis prescriptivo.

En el caso que la dirección mantenga una actitud conservadora o de alta aversión al riesgo se recomienda aplicar la optimización considerando las restricciones globales sugeridas por la herramienta. Esto es, actuar dentro de los umbrales de dos desviaciones estándar y los máximos y mínimos observados en los registros originales.

Considerando que la línea sujeta a análisis muestra típicamente una transaccionalidad de una venta por semana se recomienda en el proceso de cleansing incrementar el umbral de estabilidad máxima al 99% a los fines de no excluir el atributo de unidades vendidas del análisis y tampoco los índices de rotación del producto.

Referencias

1. Bertsimas, D., & Dunn, J. (2019). Machine learning under a modern optimization lens (D. I. LLC (ed.); 1st ed.).
2. Galkin, A. (2019). How Price Optimization Models Boost Retail Enterprises' Revenue. Competera. https://competera.net/resources/articles/price-optimization-models
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (T. Dietterich (ed.); 1st ed.). The MIT Press.

4. Sapp, C. E. (2018). Preparing and Architecting for Machine Learning. Gartner Technical Professional Advice, January, 1–38. https://doi.org/G00317328

References

1. Bertsimas, D., & Dunn, J. (2019). Machine learning under a modern optimization lens (D. I. LLC (ed.); 1st ed.).
2. Galkin, A. (2019). How Price Optimization Models Boost Retail Enterprises 'Revenue. Competera. https://competera.net/resources/articles/price-optimization-models
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (T. Dietterich (ed.); 1st ed.). The MIT Press.
4. Sapp, C. E. (2018). Preparing and Architecting for Machine Learning. Gartner Technical Professional Advice, January, 1–38. https://doi.org/G00317328

Referências

1. Bertsimas, D., & Dunn, J. (2019). Aprendizado de máquina sob uma lente de otimização moderna (D. I. LLC (ed.); 1ª ed.).
2. Galkin, A. (2019). Como os modelos de otimização de preços aumentam a receita das empresas de varejo. Competera. https://competera.net/resources/articles/price-optimization-models
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (T. Dietterich (ed.); 1ª ed.). O MIT Pressione.
4. Sapp, C.E. (2018). Preparando e arquitetando para Machine Learning. Conselho Técnico Profissional do Gartner, 1 a 38 de janeiro. https://doi.org/G00317328