UNIVERSIDAD DE ANTIOQUIA

1 8 0 3

Análisis de segmentación de clientes para la creación de estrategias comerciales

Informe presentado para optar al título de estadístico

Santiago Agudelo Vahos

Sergio Estrada Jiménez

Orientador Externo, Conozca S.A.S.
sergioej@conozca.co

Duván Humberto Cataño Salazar

Orientador Interno, Universidad de Antioquia
duvan.catano@udea.edu.co

Instituto de Matemáticas
Facultad de Ciencias Exactas y Naturales
Universidad de Antioquia

2022

1. Resumen

En el presente trabajo se muestra una aplicación de métodos de clasificación como lo son el K-means y el Random Forest a clientes que compran en tiendas del sector del retail, con el objetivo de crear campañas que estén personalizadas teniendo en cuenta el conocimiento de compra de los clientes utilizando variables relevantes como la venta, unidades, cantidad de grupos de productos, transacciones realizadas,... Por último, se muestran las diferencias entre estas segmentaciones y los resultados obtenidos en la clasificación de los clientes.

Contenido

.

2. Introducción

El marketing es generar una experiencia positiva mientras se satisface la necesidad de un mercado, creando estrategias entendiendo los hábitos de compra de las personas.

Conozca S.A.S., una empresa que tiene experiencia en el marketing relacional, dando así un servicio de calidad a empresas como Cueros Vélez, Leonisa, Centro Comercial El Tesoro, Tania, Americanino, Chevignon, Grupo Uribe, entre otros, analizando y comprendiendo el mercado con el fin de fidelizar a los clientes que compran en dichas marcas realizando estrategias para cada tipo de consumidor.

El punto más importante dentro de las prácticas realizadas en la empresa empieza con la creación de segmentación, en donde podremos identificar y agrupar a clientes que tengan comportamientos de compra (métricas) similares en cuatro grupos (AAA, A, B, C) todo esto con el fin de tener indicadores que muestren alertas o comportamientos positivos de los clientes (general y segmentados) en las marcas.

3. Marco teórico

3.1. Random Forest

El Random Forest o bosques aleatorios, es una técnica de aprendizaje supervisado basado en árboles de decisión mejorando las limitaciones que este método tiene (tendencia a sobre ajustar los datos de entrenamiento).

Lo que se hace es construir una serie de árboles de decisión en muestras de entrenamiento. Pero al construir estos árboles de decisión, se elige una muestra aleatoria de m predictores como candidatos divididos del conjunto completo de p predictores.

Esta división puede usar solo uno de esos m predictores. Se toma una muestra nueva de m predictores en cada división y elegimos $m \approx \sqrt{p}$, es decir, el número de predictores considerados en cada división es aproximado a la raíz cuadrada del número total de predictores. Esto quiere decir que cuando se construye un bosque aleatorio, en cada división del árbol, el algoritmo ni siquiera puede considerar la mayoría de los predictores.

Entonces, todos los árboles se verán bastante similares entre sí. Por lo tanto, las predicciones de los árboles estarán altamente correlacionadas. Desafortunadamente, promediar muchas cantidades altamente correlacionadas no conduce a una reducción tan grande en la varianza como promediar muchas cantidades no correlacionadas.

En problemas de clasificación, se combinar los resultados de los árboles de decisión usando soft-voting (voto suave) en donde

se le da más importancia a los resultados en la mayoría de los árboles, al combinar los resultados, unos errores se compensan con otros y tenemos una mejor predicción.

3.2. Segmentación k-means

La agrupación en clústeres de K-means es un enfoque simple y elegante para dividir un conjunto de datos en K clústeres distintos. Para realizar el agrupamiento de K- means, primero debemos especificar el número deseado de agrupamientos K, entonces el algoritmo de K- means asignará cada observación exactamente a uno de los K conglomerados que satisface las siguientes propiedades:

1. cada observación pertenece a al menos uno de los K grupos.

2. los conglomerados no se superponen, ninguna observación pertenece a más de un conglomerado.

La idea detrás del agrupamiento de K-medias es dividir las observaciones en K conglomerados de manera que la variación total dentro del conglomerado, sumada sobre todos los K conglomerados, sea lo más pequeña posible. A partir de este número k de clusters, el algoritmo coloca primero k puntos aleatorios (centroides). Luego asigna a cualquiera de esos puntos todas las muestras con las distancias más pequeñas.

A continuación, el punto se desplaza a la media de las muestras más cercanas. Esto generará una nueva asignación de muestras, ya que algunas muestras están ahora más cerca de otro

centroide. Este proceso se repite de forma iterativa y los grupos se van ajustando hasta que la asignación no cambia más moviendo los puntos. Este resultado final representa el ajuste que maximiza la distancia entre los distintos grupos y minimiza la distancia intragrupo.

4. Metodología

1. Segmento Valor:
   El Segmento valor, es una clasificación que Conozca S.A.S. realiza para segmentar a clientes entre 4 posibles grupos (AAA,A,B,C) en estos segmentos encontramos clasificados a clientes que cumplen ciertas condiciones, por ejemplo los clientes AAA contiene al mejor 10 % de la base total, los clientes pertenecientes al segmento A son los siguientes mejores 20 % de la base, el segmento B contiene al 30 % siguente de clientes y el restante 40 % de clientes entran a pertenecer al segmento C que se caracteriza por no tener tan buenas métricas con respecto a los demás segmentos.

   El orden que se les da a los clientes se da por un análisis de Random Forest en el cuál encontramos variables (tales como unidades, transacciones, monto de compra, cantidad de grupos de productos, cantidad de marcas, recencia, permanencia...) que a la hora de realizar una compra el cliente tenga un mejor hábito de compra y pueda valorar más la marca. Luego de tener las variables más importantes para la marca, se le asigna un peso a cada una de las variables

y según el comportamiento de compra de cada uno de los clientes, procedemos a asignarle uno de los 4 segmentos descritos.

Es importante recalcar que el Segmento Valor es una segmentación que se realiza mensualmente y cubre el periodo de un año, esto con el fin de tener información actualizada y garantizar que todos los clientes que estén dentro de un segmento sea ACTIVO. dentro de la marca.

Esta segmentación también es importante dentro de Conoza S.A.S. ya que permite responder requerimeintos (dada la clasificación que caracteriza a cada segmento) de una forma rápida y analizar cada grupo de clientes por aparte, dando una mejor idea del comportamiento de cada grupo de clientes.

Segmento	Clientes	%
AAA	5,698	10%
A	11,397	20%
B	17,096	30%
C	22,794	40%
Total	**56,985**	**100%**

2. Segmentación por Venta:
   En la segmentación por venta, no se hace el uso del Random Forest ya que a diferencia del Segmento Valor, se va a clasificar a los clientes en una sola variable (ventas totales en el último año por cliente).

Realizamos en RStudio una segmentación por K-means en donde se sacan en 7 grupos totales y además, se tienen en cuenta para cada uno de los grupos, los mínimos de compras, máximos de compras, un promedio de venta, todo esto con el fin de proponer a la marca varias posibles combinaciones de 3 ó 4 segmentos (TOP, ALTO, MEDIO, BAJO).

Este tipo de segmentación es mucho mejor para fidelizar clientes que el Segmento Valor, ya que es más fácil comunicarle a un cliente que para subir a una categoría mayor debe comprar x cantidad de plata. Hasta segmentación, también se actualiza mensualmente.

Grupos	Clientes	Venta	Mínimo vta	Máximo vta	Frecuencia
6	205	$ 1,220,440	$ 919,412	$ 3,522,485	5.0
5	1,597	$ 617,418	$ 493,996	$ 914,616	3.1
1	5,670	$ 369,980	$ 299,909	$ 493,529	2.3
2	13,503	$ 229,756	$ 183,950	$ 299,832	1.7
7	25,517	$ 138,138	$ 106,215	$ 183,941	1.4
3	41,275	$ 74,266	$ 49,580	$ 106,186	1.2
4	46,046	$ 24,861	-$ 396,469	$ 49,538	1.1

5. Conclusiones y Recomendaciones

Luego haber realizado las prácticas laboralescomo estadístico y analista de datos dentro de Conozca S.A.S. y haber adquirido conociminetos en el área de estadística y en otras áreas se concluye que el estudiante:

- Desarrolló conocimientos en uso del lenguage SQL, limpie-

za, gestión y análisis de bases de datos, además se adaptó a las necesidades del mercado aprendiendo el funcionamiento del sector.

- Uso de tableros de control para creación de informes mensuales y análisis requeridos por las marcas.

- Realizó distintos métodos de segmentación, entendiendo el funcionamiento de estos y aplicando conocimientos adquiridos en el pregrado de estadística.

A Conozca S.A.S. se le recomienda revisar procedimientos de pesos de las variables en los Random Forest ya que luego de el 2020, año que cambió por completo el comportamiento de venta de los clientes, no se han actualizado dichos pesos y como resultado tiene que en algunos meses existe la ausencia de los clientes del Segmento Valor C que han comprado dentro de la marca, esto porque existen clientes que llevan mucho tiempo sin compras y posiblemente los pesos estén ignorando la recencia de compra.

Es importante mencionar que el tiempo trabajando en Conozca S.A.S. ha dado un perfil prosefional relacionado al mercadeo y analítica de datos.

Referencias

[1] Aurélien Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, Sebastopol, CA, 2017.

[2] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R.* Springer, 2013.

[3] V. Mirjalili and S. Raschka. *Python Machine Learning.* Marcombo, 2020.

[4] Wes McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.* O'Reilly Media, 1 edition, February 2013.