

Perfilamiento Alertas De Gastos y Consejos Financieros

The Kernel Trick

Mejorar la experiencia del cliente con las Personal Financial Management (PFM) traerá como consecuencia natural un uso mas recurrente, fidelización y por lo tanto ganancia de participación de la app en el mercado y finalmente la buena publicidad por medio del voz a voz. Es por esto que teniendo este objetivo principal en mente el equipo The Kernel Trick a desarrollado dos herramientas analíticas para sub-segmentar a la población que utiliza PSE como medio de pago para así dar alertas de gastos y consejos financieros dirigidos y que realmente puedan interesar al cliente final, estas herramientas no deben verse como dos estrategias separadas sino por el contrario el objetivo es ensamblarlos para tener un mayor nivel de precisión.

Herramienta Analítica 1:

Se desarrolló un modelo supervisado que tiene como objetivo pronosticar la probabilidad de que una persona tenga un gasto en el futuro.

La variable a pronosticar se construyo utilizando la base de transacciones, De ella se sacó el valor transaccional que ha tenido cada uno de los clientes por mes desde un punto de observación definido 201609, así para los siguientes 12 meses. Con este vector se calculó el promedio de transacciones futuras. Finalmente, los clientes que estaban por encima del promedio del valor transaccional del total de la base se marcaron con 1 y el resto con 0. El objetivo es identificar a través de un modelo estos clientes que van a tener un alto valor transaccional en los siguientes 12 meses.

La siguiente grafica ilustra la construcción de la variable a pronosticar:

Vector De Transacciones:

	201609	t1	t2	t3	t4	...	t10	t11	t12	Avg Uti	Bin_uti
Cliente1	400000	200000	100000	200000	300000	...	300000	400000	100000	230769,2308	1
Cliente 2	0	0	0	0	0	...	0	0	0	115384,6154	0
....
Cliente N	0	0	0	0	0	...	500000	500000	500000	307692,3077	1

En la tabla anterior se calcula el promedio de transacciones de los siguientes 12 meses desde el punto de observación (Avg Uti) para cada cliente, adicionalmente, se calcula una variable binaria (Bin_uti) que toma el valor de 1 si este promedio de transacciones del cliente está por encima la mediana de toda la base, estos clientes son los que queremos identificar.

Variables Independientes:

Con la variable respuesta se construyó un *scorecard* de cada variable que se quiere utilizar en el modelo, para analizar su comportamiento:

Seg_str:

seg_str	Bin_utili	count(*)	Tasa Utilizan	Distribucion
OTRO	73	352	21%	0%
PERSONAL	90.022	209.089	43%	68%
EMPRENDEDOR	19.985	36.924	54%	12%
PERSONAL PLUS	29.819	42.875	70%	14%
PREFERENCIAL	13.547	17.632	77%	6%
Total	153.446	306.872	50%	100%

Esto quiere decir que el 77% de los clientes que están en preferencial tendrán un gasto promedio en los siguientes 12 meses mayor a \$46.000 que es la mediana de la base, mientras que los clientes que están clasificados como otros solo el 7% tendrá un comportamiento mayor al promedio de la base.

Este cuadro también nos sirve para ordenar el comportamiento de los grupos, por ejemplo, vemos como preferencial tiene gente con una mayor transaccionalidad que personal plus, emprendedor, personal y otro en ese orden.

Ocupación:

ocupación	Bin_utili	count(*)	Tasa Utilizan	Distribucion
Estudiante	8.669	27.002	32%	9%
Desempleado Con Ingresos	398	959	42%	0%
Hogar	2.240	5.224	43%	2%
Desempleado Sin Ingresos	521	1201	43%	0%
Otra	1.590	3.518	45%	1%
Comerciante	2.501	5.418	46%	2%
Empleado	96.066	191.740	50%	62%
Agricultor	117	228	51%	0%
Independiente	15.534	28.988	54%	9%
Jubilado	4.763	8.886	54%	3%
Ganadero	156	284	55%	0%
Independiente	10.223	16.570	62%	5%
Socio	5.219	8.362	62%	3%
Rentista De Capital	1.328	2.105	63%	1%
	4.120	6.383	65%	2%
Total	149.325	306.872	49%	100%

Con respecto al manejo de información faltante (nulos), se decidió para todas las variables dejarlo en una categoría aparte, porque como se ve en esta ocasión la gente que no tiene valor para esta variable puede tener un comportamiento diferenciado a los demás grupos, por ejemplo aquellos que no tienen marca de ocupación tienen una tasa de transaccionalidad el 65%.

Nota: El análisis para las siguientes variables se hizo de igual manera, en aras de no repetir y hacer más fácil la lectura no pondremos la descripción escrita.

Tipo vivienda

tipo_vivienda	Bin_utili	count(*)	Tasa Utilizan	Distribución
	60.968	152.767	40%	50%
1	1	2	50%	0%
2	3	6	50%	0%
3	1	1	100%	0%
4	0	1	0%	0%
No Informa	5.999	11.131	54%	4%
Alquilada	13.456	23.617	57%	8%
Familiar	46.930	78.019	60%	25%
Propia	26.088	41.328	63%	13%
Total	153.446	306.872	50%	100%

Nivel Académico

nivel_academico	Bin_utili	count(*)	Tasa Utilizan	Distribución
	11.076	37.733	29%	12%
Primaria	945	2.454	39%	1%
No Informa	17.244	43.545	40%	14%
Bachillerato	8.510	21.283	40%	7%
Ninguno	1.312	3.047	43%	1%
Técnico	18.719	41.034	46%	14%
Universitario	75.114	126.590	59%	42%
Postgrado	12.568	18.440	68%	6%
Especialización	6.251	8.439	74%	3%
Total	151.739	302.565	50,151%	100%

Estado Civil:

estado_civil	Bin_utili	count(*)	Tasa Utilizan	Distribución
No Informa	10.805	23.718	46%	8%
Soltero	72.581	158.666	46%	52%
Desconocido	13.616	28.791	47%	9%
Viudo	1.566	3.019	52%	1%
Divorciado	5.625	10.841	52%	4%
Otro	1.032	1.923	54%	1%
Casado	44.449	73.768	60%	24%
	3.772	6.146	61%	2%
Total	153.446	306.872	50%	100%

Genero

genero	Bin_utili	count(*)	Tasa Utilizan	Distribución
M	72.876	147.612	49%	48%
F	76.606	153.806	50%	50%
	3.964	5.454	73%	2%
Total	153.446	306.872	50%	100%

R_Ingreso

ingreso_rango	Bin_utili	count(*)	Tasa Utilizan	Distribución
a. (0 1.1MM]	7.346	37.027	20%	12%
0	190	921	21%	0%
No disponible	1.519	5.119	30%	2%
b. (1.1 2.2MM]	35.331	93.563	38%	30%
c. (2.2 3.3MM]	27.722	50.707	55%	17%
d. (3.3 4.4MM]	20.644	32.742	63%	11%
e. (4.4 5.5MM]	12.399	18.965	65%	6%
f. (5.5 6.6MM]	8.946	13.290	67%	4%
g. (6.6 7.6MM]	6.759	9.982	68%	3%
h. (7.6 8.7MM]	5.096	7.374	69%	2%
i. (8.7 Inf)	27.494	37.182	74%	12%
Total	153.446	306.872	50%	100%

Random Forest:

Con las variables categorizadas y la variable respuesta se construyó un Random Forest que tiene como resultado el siguiente ScoreCard:

Entrenamiento:

r_random	buenos	conteo	Tasa Buenos	Distribución
1	18.795	70.887	27%	33%
2	26.090	70.887	37%	33%
3	62.687	73.036	86%	34%
Total	107.572	214.810	50%	100%

KS: 48

Validación:

r_random	buenos	conteo	Tasa Buenos	Distribución
1	7.618	30.380	25%	33%
2	11.339	30.380	37%	33%
3	26.917	31.302	86%	34%
Total	45.874	92.062	50%	100%

KS: 48

Como se puede observar tanto en entrenamiento como en validación se observa un Ks similar por lo que se puede concluir que no existe overfitting.

Finalmente, con el Total De La Base se definen tres perfiles transaccionales, Alto, Medio y Bajo.

r_random	buenos	conteo	Tasa Buenos	Distribución
1	26.552	101.267	26%	33%
2	37.358	101.268	37%	33%
3	89.536	104.337	86%	34%
Total	153.446	306.872	50%	100%

KS: 48

Herramienta Analítica 2:

Por otro lado, se desarrolló una propuesta que combina modelos de aprendizaje supervisado y no supervisado. Esta solución consiste en encontrar temas (o categorías) en las cuales clasificar las transacciones. Estas categorías son encontradas mediante un modelo no supervisado que estudia los textos descriptivos de las transacciones y trata de agruparlos.

Sin embargo, menos del 30% de los datos de transacciones tienen la información completa en sus campos de texto. Por este motivo, se entrenó luego un modelo supervisado para clasificar una

transacción, a partir de datos del pagador y de la transacción, dentro de una de las categorías anteriormente determinadas.

Esta herramienta fue desarrollada en *Jupyter notebooks*, y su proceso de desarrollo está documentado en los mismos, junto al código utilizado y los resultados obtenidos. Para obtener mayor información sobre esta propuesta, por favor revisar los notebooks, que se encuentran en la carpeta **Código/Clasificación de transacciones en categorías/** de este repositorio.

Ensamble:

Las dos propuestas analíticas fueron combinadas, y al ensamblar ambos modelos es posible radicar estrategias dirigidas de alertas de gastos y consejos financieros:

Agrupación Modelo Gastos:

Otros	Otros
Trámites/gobierno	Servicios E Impuestos
Pago de impuestos	
Seguridad social	
Servicios públicos	
Transporte/construcción	hogar
Vivienda	
Tecnología y comunicaciones	
Educación	
Pago de deudas	Bancos
Tarjetas de crédito	
Ahorro y giros	Ahorro

Matriz de Intensidad

Con esta matriz se definirá con que intensidad la aplicación puede desplegar notificaciones.

Segmento Transaccional/Segmento Gasto	Servicios E Impuestos	hogar	Bancos	Ahorro	Otros
Bajo	4	3	2	1	0
Medio	3	2	1	1	0
Alto	2	1	1	1	0

Matriz de Alertas

Segmento Transaccional/Segmento Gasto	Servicios E Impuestos	hogar	Bancos	Ahorro	Otros
Bajo	Alerta Soat, Declaración, Recibos Mensuales	Impuesto Predial, Administración, Mantenimientos	Tarjetas De Crédito, créditos de consumo, créditos hipotecarios	na	na
Medio	Alerta Soat, Declaración, Recibo Bimensual	Impuesto Predial, Mantenimientos	Tarjetas De Crédito, créditos de consumo	na	na
Alto	Alerta Soat, Declaración	Impuesto Predial	Tarjetas De Crédito	na	na