

Prueba técnica
Giovanni Gamaliel López Padilla

Sección 1

Realizando un estudio sobre el comportamiento de las variables del documento dado, se encontraron las siguientes características.

- No existe una correlación fuerte contemplando todas las variables del documento con el parametro Rango score. Ya que, el valor de las correlaciones son menores a 0.2.
- Existe una mayor probabilidad que los clientes clasificados como BB abandonen el producto en comparación a los clientes clasificados como otros y AA.(Tabla 1)
- Puede que exista un sezgo en el modelo debido a que existe una diferencia clara cuando no existe dispersión en el rango disponible. Esto puede deberse a que sea un nuevo cliente. (Tabla 2)
- A mayor cantidad de productos contratados, la probabilidad de abandono se reduce.
- A mayor antigüedad, la probabilidad de abandono se reduce considerablemente.

Rango_score	Otros	AA	BB
0 % a 10 %	15.565	21.089	10.393
10.1 % a 20 %	24.272	28.649	19.464
20.1 % a 30 %	20.813	21.826	19.419
30.1 % a 40 %	16.681	14.108	16.522
40.1 % a 50 %	11.961	8.121	13.204
50.1 % a 60 %	6.73	3.94	9.591
60.1 % a 70 %	2.845	1.66	6.35
70.1 % a 80 %	1.014	0.533	3.545
80.1 % a 90 %	0.111	0.075	1.375
90.1 % a 100 %	0.009		0.138

Tabla 1: Distribución de rango score para cada clasificación.

Rango_score	Sin dispersión	0.1 a 99.9	100 a 4999.9	5000 a 9999.9	10000 a 19999.9	20000 a 39999.9	40000 a 59999.9	60000 o más
0 % a 10 %	28.467	1.299	9.053	7.228	10.672	15.87	17.97	16.554
10.1 % a 20 %	29.927	3.896	14.341	15.562	20.539	25.282	28.654	27.423
20.1 % a 30 %	24.088	6.494	14.608	17.411	20.66	22.215	22.821	23.001
30.1 % a 40 %	10.219	18.182	14.268	16.851	17.45	16.097	14.488	15.388
40.1 % a 50 %	2.19	14.286	13.495	15.086	13.503	10.502	8.445	9.359
50.1 % a 60 %	3.65	6.494	11.995	12.145	9.043	5.883	4.516	4.949
60.1 % a 70 %	0.73	12.987	10.22	8.707	5.144	2.833	2.189	2.258
70.1 % a 80 %	0.73	18.182	7.538	5.001	2.299	1.085	0.773	0.89
80.1 % a 90 %		15.584	3.95	1.848	0.653	0.223	0.145	0.175
90.1 % a 100 %		2.597	0.532	0.161	0.038	0.01		0.003

Tabla 2: Distribución de rango score para cada distribución del promedio disponible del cliente.

Rango_score	0	1	2	Más de 2
0 % a 10 %	19.048	10.696	12.487	11.369
10.1 % a 20 %	20	17.282	21.345	22.278
20.1 % a 30 %	6.667	16.616	19.82	22.131
30.1 % a 40 %	1.905	14.787	16.092	17.709
40.1 % a 50 %	4.762	13.009	12.507	12.52
50.1 % a 60 %	7.619	10.706	8.719	7.756
60.1 % a 70 %	18.095	8.253	5.446	4.148
70.1 % a 80 %	12.381	5.57	2.706	1.717
80.1 % a 90 %	9.524	2.705	0.849	0.365
90.1 % a 100 %		0.376	0.029	0.007

Tabla 3: Distribución del rango score para la cantidad de productos contratados.

Rango_score	6 o menos	7 a 12	13 a 24	25 a 60	61 a 120	121 o más
0 % a 10 %	4.855	4.812	5.608	7.796	11.248	21.27
10.1 % a 20 %	7.321	7.856	9.583	15.541	24.45	31.679
20.1 % a 30 %	7.286	10.503	13.27	19.672	24.578	22.217
30.1 % a 40 %	8.907	13.348	17.208	19.963	17.986	12.794
40.1 % a 50 %	12.127	16.64	19.056	16.601	11.205	6.652
50.1 % a 60 %	15.822	17.993	16.632	11.151	5.862	2.981
60.1 % a 70 %	18.521	16.003	11.181	5.792	2.791	1.436
70.1 % a 80 %	16.147	9.405	5.319	2.52	1.356	0.733
80.1 % a 90 %	7.951	3.107	1.932	0.894	0.501	0.23
90.1 % a 100 %	1.063	0.332	0.211	0.069	0.023	0.007

Tabla 4: Distribución del rango score para cada rango de antigüedad.

Con los puntos antes mencionados podemos delimitar a retener a las personas que tengan una antigüedad entre 7 a 12 y tengan a lo más un producto contratado. Esto debido a que pasando al rango de 13 a 24 años o que tengan al menos dos productos contratado la probabilidad de abandono baja considerablemente.

Sección 2

Identificaría las claves únicas de las bases de datos. Estas pueden ser la información de la fecha y hora. Además de una clave única de usuarios que podría ser una clave interna del banco. A cada tabla individualmente le añadiría una columna extra llamada movimiento que guardará el tipo de movimiento que es el que está registrado. Con estas claves identificadas y las columnas creadas, uniría las tablas realizando un full join, ya que no son dos grupos excluyentes debido a la naturaleza de los datos y para tener un orden las ordenaría de forma cronológica. La validación de la unión lo haría observando el número de datos que cuenta la unión de la tabla ya que deberían ser la suma de los registros de cada tabla de manera individual.

```
SELECT *
FROM credit.table AS credit
```

```
FULL OUTER JOIN sell.table AS sell  
ON credit Val = sell Val;
```

Sección 3

Emplearía un modelo basado en un sistema de recomendaciones con el objetivo de verificar la compatibilidad de un producto con el usuario, en comparación con otros usuarios que exhiben un comportamiento similar. Para alimentar este modelo, utilizaría información como el saldo disponible, ingresos y egresos del cliente en los últimos tres años. Posteriormente, compararía esta serie temporal con la de otros usuarios y analizaríamos los tipos de productos que ellos consideran. Además, tendría en cuenta el tiempo de antigüedad de otros usuarios con el producto que se desea recomendar o analizar. Como primera aproximación implementaría un modelo basado en distancias como sería la similitud coseno e ir iterando hasta llegar a un modelo robusto de imputación de datos ya que este problema puede plantearse como un problema de datos incompletos.