

客户定位分析

1 业务需求

数据集(数据来源 UCI)为 2009-2016 年商业银行部分客户的历史数据(54 个自变量, 1 个二分类因变量, 354293 条数据), 主要包括用户基本信息, 资产状况, 历史交易信息, 是否购买了该理财产品等信息, 根据现有信息, 提取有价值信息供市场决策以及预测哪些新用户会购买该理财产品.

1.1 变量说明:

序号	变量	说明
1	CUST_ID	客户号
2	GENDER	性别
3	MARR	婚姻状态
4	EDUCATION	学历
5	AGE	年龄
6	CHILDREN	子女情况
7	F_VIP	是否 VIP(惠通)
8	F_STAFF	是否本行员工
9	F_PAYROLL	是否代发
10	F_YLJ	是否代发养老金
11	F_CC	是否信用卡客户
12	F_WEB	是否签约网银
13	F_MOBILE	是否签约手机银行
14	F_TEL	是否签约电话银行
15	F_FUND	是否签约基金
16	DOB	行龄(day)
17	F_YJL	是否易精灵客户
18	F_CLOAN	是否消费贷款
19	F_HLOAN	是否房贷客户
20	AUM_3	AUM 月日均(近三月)
21	FUND_3	基金月日均(近三月)
22	DEBIT_3	国债月日均(近三月)
23	AUM_6	AUM 月日均(近六月)
24	FUND_6	基金月日均(近六月)

25	DEBIT_6	国债月日均(近六月)
26	DEPOSIT_3	存款月日均(近三月)
27	DEPOSIT_6	存款月日均(近六月)
28	FIX_3	定期存款月日均(近三月)
29	FIX_6	定期存款月日均(近六月)
30	FINACE_3	理财产品月日均(近三月)
31	FINACE_6	理财产品月日均(近六月)
32	YJL_3	易精灵月日均(近三月)
33	YJL_6	易精灵月日均(近六月)
34	PAYROLL_3	月平均代发金额(近三月)
35	PAYROLL_6	月平均代发金额(近六月)
36	C_1W_D_3	单笔 1 万元以上存款次数(近三月)
37	C_1W_D_6	单笔 1 万元以上存款次数(近六月)
38	C_1W_TR_3	单笔 1 万元以上转账入账次数(近三月)
39	C_1W_TR_6	单笔 1 万元以上转账入账次数(近六月)
40	C_FIX_3	定期存款次数(近三月)
41	C_FIX_6	定期存款次数(近六月)
42	C_FUND_3	基金交易次数(近三月)
43	C_FUND_6	基金交易次数(近六月)
44	C_YJL_3	易精灵交易次数(近三月)
45	C_YJL_6	易精灵交易次数(近六月)
46	C_DEBIT_3	国债交易次数(近三月)
47	C_DEBIT_6	国债交易次数(近六月)
48	C_FIANCE_3	理财购买次数(近三月)
49	C_FIANCE_6	理财购买次数(近六月)
50	GAP_FINACE_3	理财购买平均间隔天数(近三月)
51	GAP_FIANCE_6	理财购买平均间隔天数(近六月)
52	DT_L_FINACE	理财产品最后一次购买日期
53	CHANNEL_PRE	理财产品购买渠道偏好
54	A_L_FIANCE	理财产品最后一次购买的金额
55	Target	是否购买理财产品 1 购买,0 未购买

2. 数据描述分析(proc means)

变量	N	缺失值个数	均值	中位数	最小值	最大值
AGE	343918	10375	55.5594822	58.0000000	0	109.0000000
AUM_3	344054	10239	109910.51	6426.00	0	131770231
AUM_6	344307	9986	75748.99	5152.28	0	87610201.76
A_L_FIANCE	293306	60987	187193.21	100000.00	3000.00	130000000
CHILDREN	344468	9825	0	0	0	0
CUST_ID	354293	0	3088267210	3005067793	3000000004	6014397196
C_1W_D_3	18542	335751	1.6136339	1.0000000	1.0000000	41.0000000
C_1W_D_6	31167	323126	2.1442551	1.0000000	1.0000000	106.0000000
C_1W_TR_3	63577	290716	1.1841075	1.0000000	1.0000000	43.0000000
C_1W_TR_6	114686	239607	1.2564219	1.0000000	1.0000000	110.0000000
C_FIANCE_6	63287	291006	6.3878996	2.0000000	1.0000000	495.0000000
C_FIX_3	21214	333079	2.8991232	2.0000000	1.0000000	244.0000000
C_FIX_6	35148	319145	4.1888585	2.0000000	1.0000000	612.0000000
C_FUND_3	8282	346011	2.7751751	2.0000000	1.0000000	90.0000000
C_FUND_6	13258	341035	4.5460854	2.0000000	1.0000000	246.0000000
DEBIT_3	344054	10239	11202.55	0	0	36830000.00
DEBIT_6	344054	10239	8738.98	0	0	28996666.67
DEPOSIT_3	319194	35099	46574.20	4379.47	0.000100000	53631080.14
DEPOSIT_6	320618	33675	46019.79	5531.43	0.000100000	59214320.47
DOB	344328	9965	3272.87	3275.00	106.0000000	7797.00
DT_L_FINACE	293306	60987	19657.52	19833.00	17295.00	20635.00
FINACE_3	44683	309610	258507.64	66923.08	549.4505000	95714285.71
FINACE_6	70221	284072	231945.98	61748.63	546.4481000	106557377
FIX_3	56243	298050	156096.16	65000.00	2.1978000	47275126.53
FIX_6	61000	293293	145499.03	57907.58	0.5394000	52974466.59
FUND_3	344054	10239	14504.01	0	0	74791297.67
FUND_6	344307	9986	9522.05	0	0	49102248.60
GAP_FIANCE_6	63287	291006	99.9930001	91.0000000	0	183.0000000
PAYROLL_3	24920	329373	10397.25	5173.22	1.0000000	9500000.00
PAYROLL_6	30776	323517	11223.36	4615.50	2.0000000	9500000.00
TARGET	354293	0	0.0314994	0	0	1.0000000
YJL_6	28201	326092	364797.82	88469.95	546.4481000	74748633.88

(1) 部分变量的观测存在大量缺失

(2) 可能存在重复观测行

proc freq

TARGET	频数	百分比	累积 频数	累积 百分比
0	343133	96.85	343133	96.85
1	11160	3.15	354293	100.00

(3) 目标变量观测严重不均衡

(4) 存在不少字符型变量

(5) 变量之间存在重复

3. 数据处理

- (1) 删除重复行 (proc sql)
- (2) 数据抽样, 目标变量观测值比例应分布均衡 0:1=3:1 (proc surveyselect)
- (3) 划分训练集, 验证集 (ranuni (12345) < 0.7)
- (4) 删除缺失值过多的变量 (Excel),
- (5) 数据集中所有字符型变量重编码处理 (if...then...)
- (6) 填补缺失值, 异常值删除或替换 (max (min))
- (7) 数据归一化 (去量纲) (proc standard)
- (8) 变量进一步压缩: 相关性分析 (proc corr spearman& pearson), 主成分分析 (proc princomp), 逐步回归分析 (proc logistic)

➤ 删除重复观测行

删除除ID不同外, 其他信息均相同的冗余信息后, 剩余观测数据为 352487条数据.

➤ 抽取均衡数据集

随机抽样, 保留target=1中所有数据(11114), 抽样 target=0(33342)使1: 0占比1:3, 得到新的数据集包含44456条观测值, 55个变量.

➤ 划分训练数据集和验证数据集

训练数据 : 验证数据 = 7:3

➤ 删除缺失值占比大于85%的变量

1	变量	缺失占比	变量	缺失占比
2	C_DEBIT_3	99.54%	DT_L_FINACE	12.15%
3	C_DEBIT_6	99.17%	CHANNEL_PRE	12.15%
4	C_FUND_3	96.75%	DEPOSIT_3	7.41%
5	C_YJL_3	95.50%	DEPOSIT_6	7.07%
6	C_FUND_6	94.90%	AGE	2.19%
7	YJL_3	94.31%	AUM_3	2.16%
8	C_1W_D_3	92.98%	DEBIT_3	2.16%
9	PAYROLL_3	92.95%	DEBIT_6	2.16%
10	C_FIX_3	92.08%	FUND_3	2.16%
11	C_YJL_6	92.02%	AUM_6	2.06%
12	YJL_6	91.37%	FUND_6	2.06%
13	PAYROLL_6	91.14%	CHILDREN	2.01%
14	C_1W_D_6	88.83%	DOB	1.92%
15	EDUCATION	88.32%	GENDER	0.03%
16	C_FIX_6	87.50%	CUST_ID	0.00%
17	MARR	85.04%	TARGET	0.00%
18	C_FIANCE_3	81.77%	F_CC	0.00%
19	FIX_3	79.95%	F_CLOAN	0.00%
20	C_1W_TR_3	79.63%	F_FUND	0.00%
21	FIX_6	78.33%	F_HLOAN	0.00%
22	FINACE_3	77.12%	F_MOBILE	0.00%
23	C_FIANCE_6	70.75%	F_PAYROLL	0.00%
24	GAP_FIANCE_6	70.75%	F_STAFF	0.00%
25	FINACE_6	68.40%	F_TEL	0.00%
26	C_1W_TR_6	63.20%	F_VIP	0.00%
27	A_L_FIANCE	12.15%	F_WEB	0.00%
28	GAP_FINACE_3	0.00%	F_YJL	0.00%
29			F_YLJ	0.00%

	变量数
删除前	55
删除	16
删除后	39
需要进一步压缩的变量数	37

➤ 字符型变量重编码

F_VIP(VIP) GENDER(性别) F_TEL(签约电话银行)

GAP_FINACE_3(近三个月购买理财产品间隔) F_STAFF(本行员

工) C_FIANCE_3(近三个月理财产品购买次数) F_PAYROLL(代

发工资) CHANNEL_PRE(该理财产品购买渠道) F_MOBILE(签约手机银行) F_YLJ(代发养老金) F_HLOAN(房贷) F_YJL(YLJ用户) F_FUND(签约基金) F_WEB(签约网银) F_CLOAN(消费贷款) F_CC(信用卡用户)

得到纯数值型变量的训练数据集data_train, 验证数据集data_test.

➤ 填补缺失值, 异常值删除或替换

连续变量缺失值填补: 中位数填补

分类变量缺失值填补: 众数填补, 备注: C_FIANCE_3(近三个月理财产品购买次数) 将缺失值单独做一类填充.

离群值替换: 超出3sigma范围的数据视为离群值, 使用中位数替换;

➤ 去量纲(数据归一化)

$$X = \frac{Xi - \mu}{\sigma}$$

➤ 根据相关性进一步删除变量

连续变量之间的相关性 (Pearson)

Pearson 相关系数, N = 45568																							
Prob > r under H0: Rho=0																							
	FIX_3	FIX_6	FINACE_3	FINACE_6	C_FIANCE_3	C_FIANCE_6	GAP_FIANCE_3	GAP_FIANCE_6	C_1W_T_3	C_1W_T_6	DEPOSIT_3	DEPOSIT_6	AUM_3	AUM_6	DEBIT_3	DEBIT_6	FUND_3	FUND_6	AGE	DOB	A_L_FIANCE_NCE	DT_L_FIANCE_NCE	
FIX_3	1	0.86288	0.00207	0.00501	-0.01708	-0.02271	0.016	0.02485	-0.01633	-0.0352	0.1313	0.13189	0.08855	0.08532	0.05121	0.05143	-0.03159	-0.03222	0.02619	0.02216	0.03389	-0.01025	
FIX_6	0.86288	1	-0.00073	0.00454	-0.01679	-0.02258	0.01698	0.0263	-0.02342	-0.04168	0.1539	0.14694	0.10033	0.09812	0.05508	0.0553	-0.02924	-0.02944	0.03134	0.03192	0.02814	-0.01202	
FINACE_3	0.00207	-0.00073	1	0.56978	0.15245	0.36417	-0.1638	-0.33471	0.02188	0.02687	0.02741	0.02868	0.07356	0.06434	-0.01504	-0.01485	0.01818	0.02034	-0.02353	-0.02205	0.21522	-0.01309	
FINACE_6	0.00501	0.00454	0.56978	1	0.1108	0.38254	-0.12007	-0.34583	0.01443	0.0458	0.0656	0.0459	0.12202	0.12519	-0.01486	-0.01486	0.0275	0.02895	-0.02269	-0.02134	0.25511	-0.0039	
C_FIANCE_3	-0.01708	-0.01679	0.15245	0.1108	1	0.48868	-0.07485	-0.43862	0.0639	0.07037	-0.0247	-0.02147	0.00084	0.00012	-0.00158	-0.00169	0.0481	0.04822	-0.03187	-0.04899	0.06402	-0.06382	
C_FIANCE_6	-0.02271	-0.02258	0.36417	0.38254	0.48868	1	-0.52031	-0.00331	0.05922	0.10015	-0.04758	-0.04686	0.01966	0.02125	-0.00386	-0.00383	0.04255	0.04358	-0.05632	-0.04956	0.08397	-0.10032	
GAP_FIANCE_3	0.016	0.01698	-0.1638	-0.12007	-0.07485	-0.52031	1	0.47549	-0.06576	-0.07829	0.01458	0.01245	-0.01789	-0.01699	-0.00034	-0.00024	-0.05009	-0.05059	0.03698	0.05239	-0.08806	0.05375	
GAP_FIANCE_6	0.02485	0.0263	-0.33471	-0.34583	-0.43862	-0.00331	0.47549	1	-0.05391	-0.10398	0.00559	0.00684	-0.05946	-0.05821	0.00216	0.00232	-0.05788	-0.05919	0.06374	0.05937	-0.09385	0.04747	
C_1W_T_3	-0.01633	-0.02342	0.02188	0.01443	0.0639	0.05922	-0.06576	-0.05391	1	0.57998	0.01229	0.01031	0.00448	0.00318	0.02554	0.02538	0.05814	0.05684	-0.01325	0.03981	0.01646	0.00054	
C_1W_T_6	-0.02342	-0.04168	0.03267	0.0458	0.07037	0.10015	-0.07829	-0.10398	0.57998	1	0.0512	0.04153	0.0484	0.03921	0.03786	0.03752	0.08191	0.08244	-0.01102	0.0686	0.03045	0.01959	
DEPOSIT_3	0.1313	0.1339	0.02741	0.0656	-0.0247	-0.04758	0.01458	0.00559	0.01229	0.0512	1	0.95033	0.79787	0.78275	0.10609	0.10618	0.09606	0.098	0.08281	0.25423	0.02582	0.26952	
DEPOSIT_6	0.13189	0.14694	0.02868	0.0459	-0.02147	-0.04686	0.01245	0.00684	0.01031	0.04153	0.95033	1	0.77387	0.78558	0.10902	0.10916	0.09897	0.10127	0.07707	0.24181	0.03238	0.28972	
AUM_3	0.08855	0.10033	0.07356	0.12202	0.00064	0.01966	-0.01789	-0.05846	0.00448	0.0484	0.79787	0.77387	1	0.8812	0.21797	0.21781	0.2354	0.23642	0.08988	0.24763	0.05689	0.33201	
AUM_6	0.08532	0.09812	0.06434	0.12519	0.00012	0.02125	-0.01699	-0.05821	0.00318	0.03921	0.78275	0.78558	0.8812	1	0.21493	0.21496	0.23284	0.23575	0.0854	0.23236	0.06071	0.34432	
DEBIT_3	0.05121	0.05508	-0.01504	-0.01486	-0.00158	-0.00386	-0.00034	0.00216	0.02554	0.03786	0.10609	0.10902	0.21797	0.21493	1	0.9944	0.05745	0.05687	0.07367	0.10113	-0.03782	0.05054	
DEBIT_6	0.05143	0.0553	-0.01485	-0.01486	-0.00169	-0.00383	-0.00024	0.00232	0.02538	0.03752	0.10618	0.10916	0.21781	0.21496	0.9944	1	0.05731	0.05674	0.07378	0.10117	-0.03772	0.05054	
FUND_3	-0.03159	-0.02924	0.01918	0.0275	0.0481	0.04255	-0.05009	-0.05788	0.05814	0.09191	0.06006	0.09897	0.2354	0.23284	0.05745	0.05731	1	0.9894	-0.0043	0.0448	0.02262	0.08867	
FUND_6	-0.02322	-0.02044	0.02034	0.02895	0.04822	0.04358	-0.05059	-0.05919	0.05884	0.09244	0.098	0.10127	0.23642	0.23575	0.05687	0.05674	0.9894	1	-0.00414	0.04636	0.02256	0.09041	
AGE	0.02619	0.03134	-0.02353	-0.02269	-0.03187	-0.05632	0.03698	0.06374	-0.01325	-0.01102	0.08281	0.07707	0.08988	0.0854	0.07367	0.07378	-0.0043	-0.00414	1	0.27883	-0.08878	-0.02051	
DOB	0.02216	0.03192	-0.02305	-0.02134	-0.04899	-0.04956	0.05239	0.05937	0.03981	0.0686	0.25423	0.24181	0.24763	0.23236	0.10113	0.10117	0.0448	0.04636	0.27883	1	-0.05956	-0.04863	
A_L_FIANCE	0.03389	0.02814	0.21522	0.25511	0.06402	0.08397	-0.08806	-0.09385	0.01646	0.03045	0.02582	0.02238	0.05689	0.06071	-0.03782	-0.03772	0.02262	0.02256	-0.08878	-0.05956	1	0.04863	
DT_L_FIANCE	-0.01025	-0.01202	-0.01309	-0.0039	-0.06382	-0.10032	0.05375	0.04747	0.00054	0.01959	0.26952	0.28972	0.33201	0.34432	0.05054	0.05054	0.08867	0.09041	-0.02051	-0.00405	0.04863	1	

连续变量与分类变量之间的相关性（Spearman）

Spearman 相关系数. N = 45568															
	TARGET	F_CC	F_CLOA N	F_FUND	F_HLOA N	F_MOBI LE	F_PAYR OLL	F_STAF F	F_TEL	F_VIP	F_WEB	F_YJL	F_YLJ	CHANN EL_PRE	GENDE R
FIX_3	0.03428	-0.03011	-0.02058	-0.02194	-0.01011	-0.03793	-0.01242	-0.04022	-0.01082	0.08873	-0.03348	-0.02186	0.04186	-0.03477	-0.00002
FIX_6	0.03128	-0.0334	-0.02151	-0.02295	-0.01364	-0.04	-0.01434	-0.04573	-0.0097	0.0823	-0.0347	-0.03151	0.04625	-0.03814	-0.00246
FINACE_3	0.02585	0.02512	0.01446	0.01155	0.00433	0.00786	-0.01261	0.00381	0.02711	0.15882	0.01427	0.18098	-0.02475	0.005	0.01391
FINACE_6	0.03307	0.02412	0.00471	0.0162	0.0106	0.00419	-0.02121	-0.00442	0.03131	0.19516	0.00643	0.2514	-0.03102	-0.00002	0.01278
C_FIANCE_3	-0.16892	0.04061	0.0308	0.01804	0.01552	0.03724	0.01359	0.04131	0.00946	0.09127	-0.01747	0.25062	-0.0607	0.01217	0.01581
C_FIANCE_6	-0.20754	0.05028	0.03494	0.01257	0.0168	0.03574	-0.00004	0.05249	0.02456	0.10167	-0.05507	0.29317	-0.08446	-0.00258	0.01813
GAP_FINACE_3	0.16794	-0.04634	-0.03415	-0.01994	-0.01843	-0.04471	-0.01603	-0.04793	-0.01305	-0.10045	0.01309	-0.28332	0.06277	-0.0179	-0.01626
GAP_FIANCE_6	0.0854	-0.06421	-0.0347	-0.03494	-0.01289	-0.07024	-0.00858	-0.05142	-0.03464	-0.13718	-0.00876	-0.35705	0.06502	-0.0447	-0.00699
C_1W_TR_3	-0.05784	-0.06228	0.01204	-0.03297	-0.01031	-0.11117	-0.04684	0.0378	-0.04439	-0.1075	-0.107	-0.18225	-0.05982	-0.07219	-0.0002
C_1W_TR_6	-0.05943	-0.01861	0.05331	-0.00867	0.01161	-0.06667	-0.00659	0.10561	-0.03188	-0.05261	-0.10143	-0.0904	-0.08708	-0.05222	0.00759
DEPOSIT_3	0.31791	0.08069	0.02889	0.08406	0.02319	0.0328	0.13388	0.06131	0.07076	0.07667	0.1296	0.09426	0.37699	0.00543	-0.0536
DEPOSIT_6	0.31088	0.08374	0.02754	0.08623	0.02216	0.04114	0.13257	0.05852	0.0746	0.09404	0.14051	0.09394	0.36799	0.0136	-0.05254
AUM_3	0.37203	0.10802	0.03503	0.23522	0.02278	0.08063	0.13727	0.06371	0.10572	0.20665	0.21149	0.27402	0.34074	0.00997	-0.05933
AUM_6	0.37128	0.10992	0.03398	0.23377	0.02114	0.09061	0.13416	0.0616	0.10571	0.22396	0.22017	0.28219	0.32929	0.02145	-0.05593
DEBIT_3	0.02991	-0.03302	-0.00889	0.0193	-0.02143	-0.0689	-0.02018	0.01236	-0.05271	0.00305	-0.15401	-0.01741	0.03	0.08961	-0.03339
DEBIT_6	0.02997	-0.03315	-0.00893	0.0192	-0.02146	-0.069	-0.02029	0.01228	-0.05259	0.00331	-0.15381	-0.01748	0.03022	0.0893	-0.0334
FUND_3	0.05079	0.06917	0.05404	0.82416	0.0127	0.06999	0.0544	0.11861	0.00524	0.04207	-0.02998	0.09627	-0.01206	0.12497	-0.04454
FUND_6	0.05324	0.07111	0.05176	0.81278	0.01198	0.07517	0.05704	0.11803	0.00891	0.04433	-0.02197	0.09772	-0.00799	0.12026	-0.04349
AGE	0.08015	-0.17333	-0.08362	-0.00124	-0.09309	-0.28574	-0.1093	-0.11968	-0.04545	-0.11183	-0.17899	-0.08264	0.48135	-0.25374	0.10742
DOB	0.09607	0.04709	0.00919	0.04971	0.06016	-0.12097	0.09909	0.02662	0.10206	-0.05784	-0.01914	-0.06516	0.52752	-0.17902	-0.0519
A_L_FIANCE	-0.00758	0.05015	0.02842	-0.00138	0.03814	0.02885	-0.01557	0.00491	0.04997	0.41785	0.05219	0.13203	-0.12108	0.09518	0.03927
DT_L_FIANCE	0.72464	0.06671	-0.0099	0.10934	-0.02858	0.14354	0.06184	0.00557	0.06986	0.07124	0.33391	0.07528	0.12606	0.30836	-0.06351

分类变量之间的相关性（VIF检验）

参数估计		参数估计							
变量	方差 膨胀	变量	方差 膨胀						
Intercept	0.00	AGE	1.23						
FIX_3	3.94	DOB	1.50						
FIX_6	3.97	A_L_FIANCE	1.31						
FINACE_3	1.58	DT_L_FIANCE	1.47						
FINACE_6	1.71	F_CC	1.16						
C_FIANCE_3	20.71	F_CLOA N	1.19						
C_FIANCE_6	6.40	F_FUND	6.62						
GAP_FINACE_3	21.95	F_HLOA N	1.06						
GAP_FIANCE_6	5.60	F_MOBI LE	1.22						
C_1W_TR_3	1.52	F_PAYR OLL	1.19						
C_1W_TR_6	1.61	F_STAF F	1.37						
DEPOSIT_3	14.12	F_TEL	1.19						
DEPOSIT_6	13.19	F_VIP	1.50						
AUM_3	35.44	F_WEB	1.49						
AUM_6	34.48	F_YJL	1.77						
DEBIT_3	895.35	F_YLJ	1.70						
DEBIT_6	895.24	CHANN EL_PRE	1.25						
FUND_3	54.84	GENDE R	1.03						
FUND_6	47.88								

VIF>=10 具有较强共线性的变量有：
C_FIANCE_3，GAP_FINACE_3，
DEPOSIT_3，DEPOSIT_6，
AUM_3，AUM_6，DEBIT_3，
DEBIT_6，FUND_3，FUND_6

删除共线性变量(13)：FIX_3 C_FINACE_3 C_FINACE_6

GAP_FINACE_3 DEPOSIT_3 DEPOSIT_6 AUM_3 DEBIT_3 DT_L_FIANCE

FUND_3 FUND_6 FINACE_3 C_1W_TR_3

剩余连续变量(9)：FIX_6 FINACE_6 GAP_FIANCE_6 C_1W_TR_6

AUM_6 DEBIT_6 AGE DOB A_L_FIANCE,

剩余分类变量(14): F_CC F_CLOAN F_FUND F_HLOAN F_MOBILE
F_PAYROLL F_STAFF F_TEL F_VIP F_WEB F_YJL F_YLJ CHANNEL_PRE
GENDER.

➤ 逐步回归方法进一步压缩变量

变量逐一进入模型方程，在引入变量时利用残差平方，当方程加入了该变量后，一旦该变量对残差的变化不显著时即可删除该变量，如此下去，逐一删除变量。

压缩后剩余显著变量有(7):

AUM_6	AUM月日均(近六月)
AGE	年龄
GAP_FIANCE_6	理财购买平均间隔天数(近六月)
CHANNEL_PRE	理财产品购买渠道偏好
F_WEB	是否签约网银
DOB	行龄(day)
GENDER	性别

4. 回归模型预测

```
ods graphics on;  
  
proc logistic data=bank.data_train_std  
  
outmodel=bank.data_model plots(MAXPOINTS=5000)=roc;  
  
class CHANNEL_PRE F_WEB GENDER;  
  
model target(event='1')=&var1 /pprob=0.40 ctable; *参数可调  
  
score out=bank.data_train_score;  
  
run;
```



```
proc logistic data=bank.data_test_std

plots(MAXPOINTS=5000)=roc;

class CHANNEL_PRE F_WEB GENDER;

model target(event='1')=&var2;

run;

ods graphics off;
```

最大似然估计分析						
参数		自由度	估计	标准 误差	Wald 卡方	Pr > 卡方
Intercept		1	-6.2444	0.2781	504.2377	<.0001
AUM_6		1	0.5920	0.0137	1865.4159	<.0001
CHANNEL_PRE	0	1	0.0387	0.0455	0.7225	0.3953
CHANNEL_PRE	1	1	0.7974	0.0439	329.5257	<.0001
CHANNEL_PRE	2	1	1.8158	0.0644	795.7958	<.0001
F_WEB	0	1	-0.1222	0.0256	22.7960	<.0001
DOB		1	-0.3675	0.0570	41.4935	<.0001
AGE		1	2.0311	0.1548	172.2645	<.0001
GENDER	0	1	0.1005	0.0170	34.9795	<.0001
GAP_FIANCE_6		1	0.0480	0.00750	40.9807	<.0001

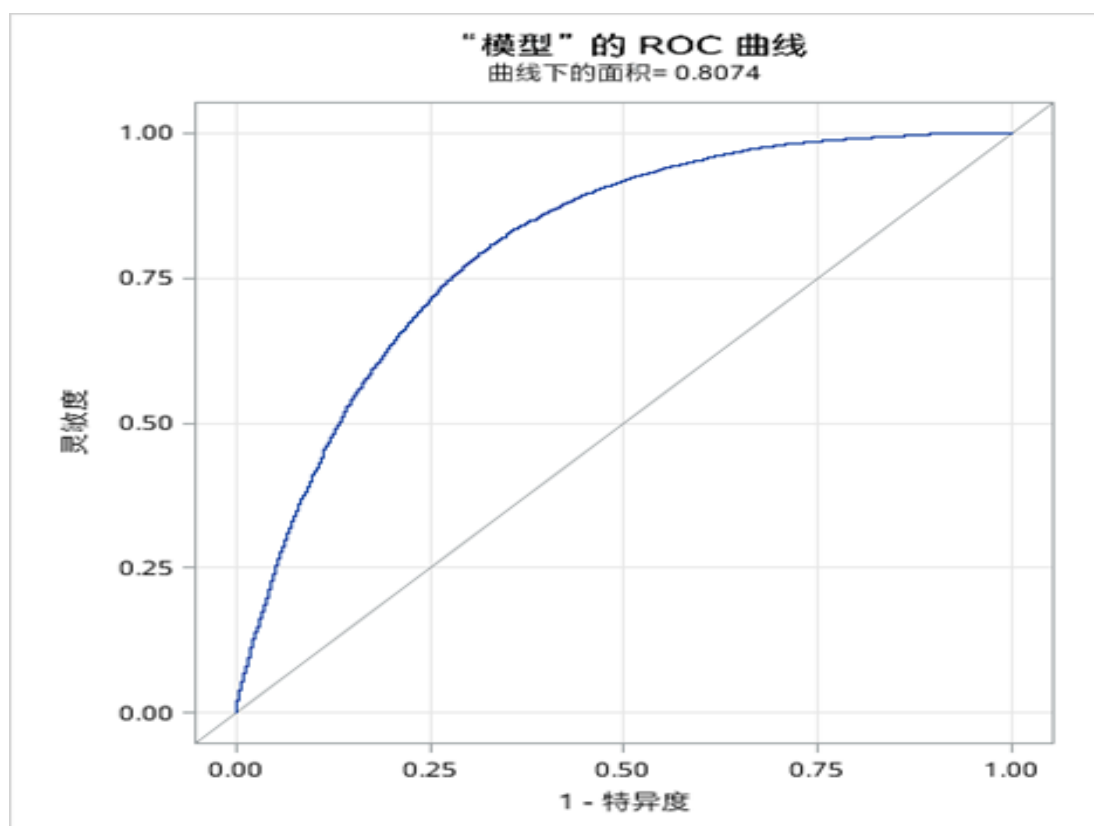
模型回归方程 $Z = 0.5920 \times \text{AUM_6} + 0.8884 \times \text{CHANNEL_PRE} -$

$0.1222 \times \text{F_WEB} -$

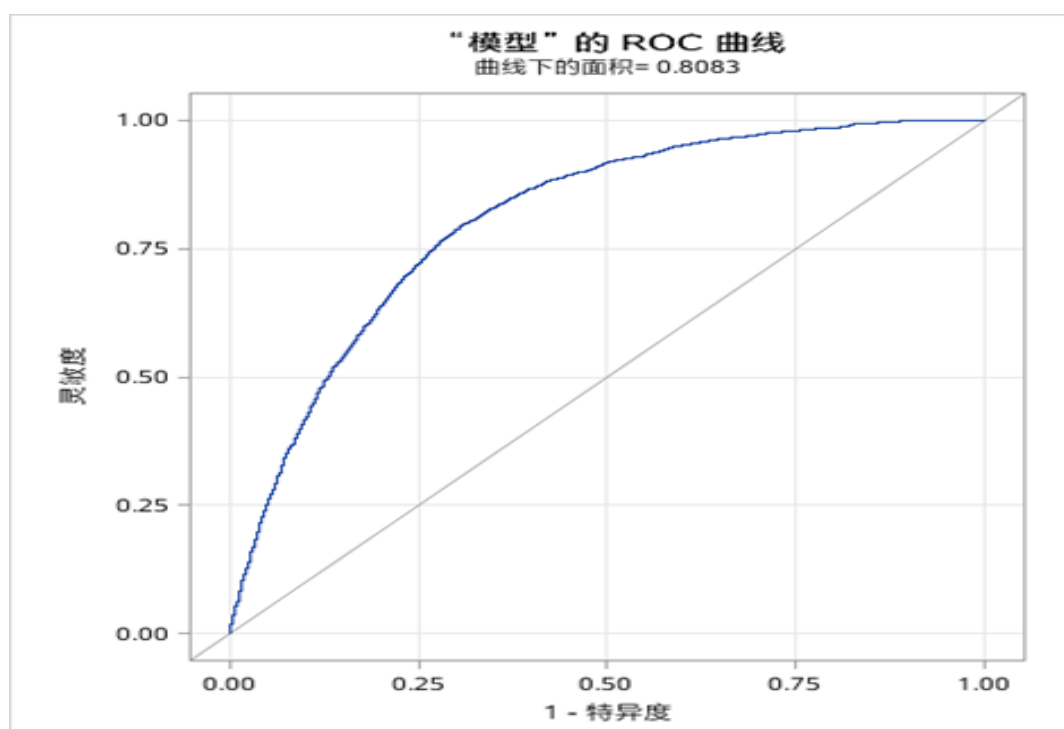
$0.3675 \times \text{DOB} + 2.0311 \times \text{AGE} + 0.1005 \times \text{GENDER} + 0.0480 \times \text{GAP_FINACE_6} -$

6.2444

$$y = \frac{1}{1+e^z}$$



训练集准确率=0.8074



验证集准确率=0.8083

5 分析与结论

- (1) 多关注年龄偏高的用户群体
- (2) 引导用户使用网络渠道
- (3) 关注行龄比较久的客户, 这类客户购买意向较高
- (4) 在历史购买人群中, 女性占多数, 因此着重关注女性用户

预测功能

如果要向新用户推荐该理财产品, 可着重关注近六月在我行的月均流水, 该用户的年龄, 之前该买别的理财产品的频率, 购买渠道偏好, 是否已开通网银及手机银行, 账龄, 性别等维度信息, 通过模型进行预测该用户会够买该理财产品的概率. 从而让理财产品经理在筛选目标客户有所放矢, 提高工作效率及成功营销概率.