1. 存量用户流失预测模型

选取因素：

CUST\_ITEM 用户属性:0,1

PREPAY\_FLAG 付费方式：0-9，12-15

USER\_STATUS 用户状态：0，1，2，4-9，A,B,e,F,g,G,I,J,j,M,N,P,Q,R,S,T,V

CREDIT\_CLASS 信用等级：连续变量

basic\_credit\_value 基本信用度：连续变量

CREDIT\_VALUE 信用额度： 连续变量

SCORE\_VALUE 累计积分： 连续变量

SEX ：NULL，0,1,F,M,m,O（缺失值较多）

USER\_FULL 资料是否完整：1（均不完整）

is\_phone 是否有联系电话： 0，1

is\_newm 是否当日新入网用户：0，1

is\_23to4 是否23g转4g：0，1

is\_hong 是否红名单用户：0，1

zhufu 是否主副卡：1，2，3

is\_mwu 是否当月三无：0，1

zhutime\_d3 前3日主叫通话时长：连续变量

gprsdata\_d3 前3日流量：连续变量

smscnt\_d3 前3日短信：连续变量

call\_10086 连续3日呼叫10086次数：连续变量

call\_10000 连续3日呼叫10000次数：连续变量

zhu\_time 当月主叫通话时长：连续变量

bei\_time 当月被叫通话时长：连续变量

zhu\_timea3 连续3日主叫通话时长：连续变量

call\_fee 当月通话费用：连续变量

zhu\_cnta 当月拔打友商客服次数：连续变量

zhu\_cnt当月主叫次数：连续变量

bei\_cnta当月接听友商客服次数：连续变量

gprs\_data当月流量：连续变量

zhuday无主叫天数：连续变量

beiday无被叫天数：连续变量

dataday无流量天数：连续变量

smsday无短信天数：连续变量

up\_sms当月短信条数（主发）：连续变量

sp\_cnt当月增值次数：连续变量

sp\_fee当月增值费用：连续变量

hu\_cnta当月呼转友商次数：连续变量

hucount连续呼转天数：连续变量

hubcount连续呼转异网天数：连续变量

结论：

付费方式：相对于第0种付费方式，2、3、5、9、12、13、14、15降低用户离网率、1、4、6、7增加用户离网率

主副卡，相对于第1种方式，第2种会降低离网率、第三种增加离网率

用户属性：1的用户离网率更低

入网天数越多的用户离网率越低（显著）

基本信用度越高离网率越高

信用额度越高离网率越低

累积积分越高离网率越低（显著）

2、3转4、有联系电话、有红名单、新入网用户离网率低

前3日主叫通话时长越高的用户，离网率越低（显著）

前3日流量越高的用户，离网率越高（显著）

前3日短信次数越高的用户离网率越高

连续3日呼叫10086次数越高离网率越低

连续3日呼叫10000次数越高离网率越低

当月主叫通话时长越长，离网率越低（显著）

当月被叫通话时长越长，离网率越低（显著）

连续3日主叫通话时长，离网率越低（显著）

当月通话费用越高，离网率越低

拨打友商客服次数越高，离网率越低

当月主叫次数越高，离网率越低（显著）

当月接听友商客服次数越高，离网率越低

当月流量使用越多，离网率越高（显著）

当月无主叫天数越多，离网率越低（显著）

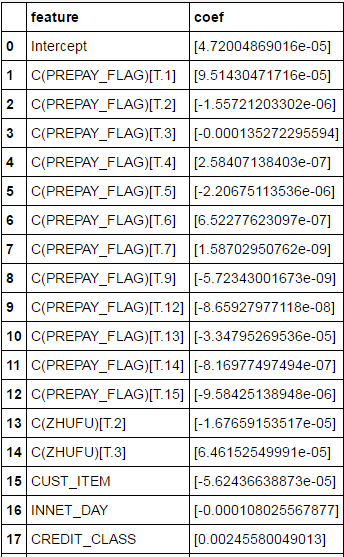
当月无被叫天数越多，离网率越低

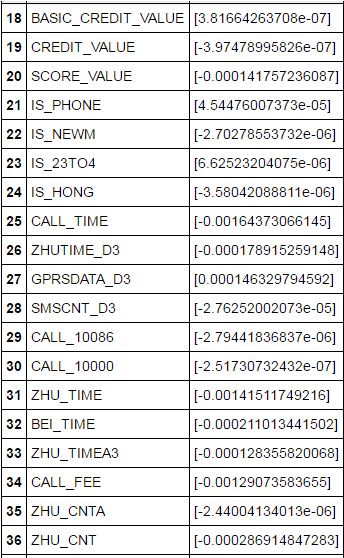
当月无流量天数越多，离网率越低

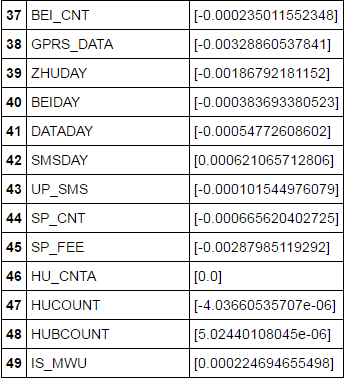
当月无短信天数越多，离网率越高

当月增值费用越高，离网率越低（显著）

模型回归系数：







100W数据结论：0.25 SIZE

precision recall f1-score support

0.0 0.88 0.99 0.93 213997

1.0 0.76 0.21 0.33 36003

avg / total 0.86 0.88 0.85 250000