

# 课程大作业：信用卡意向预测问题

## 题目背景：

GAMMA银行是一家私人银行，经营各种银行产品，如储蓄账户、活期账户、投资产品、信贷产品等。该行还向现有客户交叉销售产品，为此，客户使用不同的通信方式，如电视广播、电子邮件、网上银行推荐、手机银行等。在这种情况下，GAMMA客户银行希望将其信用卡交叉销售给现有客户。银行已经确定了一组有资格使用这些信用卡的客户。

现在，银行正在寻求您的帮助，以确定可能对推荐的信用卡表现出更高意向的客户。

## 分析目标：

通过银行收集到的客户属性数据，预测客户是否对当前推出的信用卡感兴趣。

## 数据样例：

ID	Gender	Age	Region_code	Occupation	Channel_Code	Vintage	Credit_Product	Avg_Account_Balance	Is_Active	Is_Lead
NNVBBKZB	Female	73	RG268	Other	X3	43	No	1045696	No	0
IDD6ZUNG	Female	30	RG277	Salaried	X1	32	No	581988	No	0

## 各字段含义：

Variable	Definition
ID	ID for Customer
Gender	Gender of Customer
Age	Age of Customer
Region_Code	Code of the Region for the Customer
Occupation	Occupation Type for the Customer
Channel_Code	Acquisition Channel Code for the Customer(Encoded)
Vintage	Vintage for the Customer(In Months)
Credit_Product	If the Customer has any active credit product
AvgAccountBalance	Average Account Balance for the Customer in last 12 Months
Is_Active	If the Customer is Active in last 3 Months
Is_Lead(Target)	If the Customer is interested for the Credit Card 0:not interested 1:interested

```
0    150003
1     46577
Name: Is_Lead, dtype: int64
```

RG268	28715
RG283	23620
RG254	21334
RG284	15461
RG277	10281
RG280	10232
RG269	6264
RG270	6104
RG261	6073
RG257	4910
RG251	4732
RG282	4682
RG272	4225
RG274	4218
RG281	4063
RG273	3593
RG252	3493
RG279	3233
RG263	2983
RG275	2597
RG260	2502
RG256	2258
RG276	2235
RG264	2219
RG259	2087
RG250	1991
RG255	1614
RG258	1538
RG253	1502
RG278	1437
RG262	1429
RG266	1269
RG271	1241
RG265	1239
RG267	1206

Name: Region\_Code, dtype: int64

X1	82862
X3	55064
X2	54180
X4	4474

Name: Channel\_Code, dtype: int64

Self_Employed	80732
Salaried	57486
Other	56252
Entrepreneur	2110

Name: Occupation, dtype: int64

*Is\_Lead* 即为需要预测的内容

训练集：196580

测试集：49145(without *Is\_Lead*)

## 评价指标:

**AUC**(Area Under Curve) : ROC曲线下与坐标轴围成的面积。AUC的取值范围在0.5和1之间。AUC越接近1, 检测方法的真实性越高。

```
from sklearn.metrics import roc_curve, auc
fpr, tpr, thresholds = roc_curve()
roc_auc = auc(fpr,tpr)
```

## 作业要求:

### 分组要求:

两人一组, 一组可三人 (31+28), 可跨班组队。周末前将组队结果报告给学委。

### 实验结果:

**提交方式:** 在11.8号以后, 我们会开放提交通道排行榜网页, 每组提交次数最多三次。文件命名要求后面会再说。

**文件格式:** 每行对应测试集每行数据的预测结果

Is_Lead
0.930
0.015
0.873

输出的预测值0-1之间

### 最终完成作业:

**提交内容:** 代码 作业报告 最终预测结果文件 排行榜结果排名截图 组内成员分工明细。

{报告包含: 问题分析, 数据处理流程, 选取何种方法以及原因 (若选取多种方法说明对比) }

{最终预测结果要与排行榜的值对应, 建议大家每次提交时保留原来文件}

**命名格式:** 压缩包以组号命名即可, 例如“1组.zip”

**评分标准:** 综合考虑代码、报告质量和排名, 在作业提交截止以后, 排名前十五名的组可以自愿报名参加大作业答辩 (就是讲解自己的作业), 也会加分。

**截至提交时间:** 11.14

