# 作业三——客户星级和信用等级评估

## 团队成员分工

| 姓名 | 主要工作 |
| --- | --- |
| 郑伟鑫 | star\_level, credit\_level部分的数据处理和训练模型并预测，文档编写 |
| 王文渊 | star\_level模型训练 |
| 谢瀚杵 | star\_level部分的数据处理 |
| 徐晨阳 | credit\_level部分的数据处理 |

## 数据处理的思路和处理过程

### star\_level部分

* 由于star\_level与交易数据与客户存款有关，本小组将每个用户的每个类型的交易总额提取出来并进行预测
* 使用到的数据处理过程由以下代码给出

create table   
sum\_etc ENGINE=Memory as   
select uid, sum(tran\_amt\_fen) as etc  
from dm\_v\_tr\_etc\_mx  
group by uid

create table   
sum\_grwy ENGINE=Memory as   
select uid, sum(tran\_amt) as grwy  
from dm\_v\_tr\_grwy\_mx  
group by uid

create table   
sum\_sa ENGINE=Memory as   
select uid, sum(tran\_amt) as sa  
from dm\_v\_tr\_sa\_mx  
group by uid

create table   
sum\_sbyb ENGINE=Memory as   
select uid, sum(tran\_amt\_fen) as sbyb  
from dm\_v\_tr\_sbyb\_mx  
group by uid

create table   
sum\_sdrq ENGINE=Memory as   
select uid, sum(tran\_amt\_fen) as sdrq  
from dm\_v\_tr\_sdrq\_mx  
group by uid

create table   
sum\_sjyh ENGINE=Memory as   
select uid, sum(tran\_amt) as sjyh  
from dm\_v\_tr\_sjyh\_mx  
group by uid

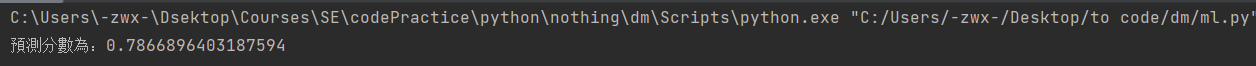
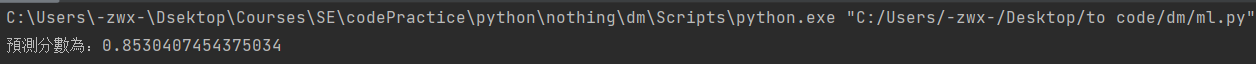
create table   
sum\_asset ENGINE=Memory as   
select uid, max(all\_bal) as asset   
from pri\_cust\_asset\_info  
group by uid

select a.uid as uid, asset, etc as etc, grwy as grwy, sa as sa, sbyb as sbyb, sdrq as sdrq, sjyh as sjyh, star\_level  
from pri\_star\_info a  
left join sum\_asset b on a.uid=b.uid  
left join sum\_etc c on a.uid=c.uid  
left join sum\_grwy d on a.uid=d.uid  
left join sum\_sa e on a.uid=e.uid  
left join sum\_sbyb f on a.uid=f.uid  
left join sum\_sdrq g on a.uid=g.uid  
left join sum\_sjyh h on a.uid=h.uid

### credit\_level部分

* 由于credit\_level与交易数据与客户贷款贷记业务有关，本小组将表pri\_cust\_liab\_info中各项数据提取用于模型训练
* 主要的数据处理逻辑如下述代码
* select  
   c.uid uid,  
   l.all\_bal all\_bal,  
   l.bad\_bal bad\_bal,  
   l.due\_intr due\_intr,  
   l.norm\_bal norm\_bal,  
   l.delay\_bal delay\_bal,  
   c.credit\_level credit\_level  
  from pri\_credit\_info c  
  left join pri\_cust\_liab\_info l on c.uid=l.uid   
  where c.credit\_level!='-1'

### 模型训练

* 本次实验，小组基于python和sklearn框架，使用KNN模型进行预测，具体的实现细节可见源码
  + 我们将此前获得的数据，分为训练集，测试集和预测集
  + 训练集与测试集的比例为7:3
* 最终训练出的模型的准确率如下图所示
  + star\_level部分：
    - 
  + credit\_level部分：
    - 
* 预测结果分别保存在predict\_star\_level.txt,predict\_credit\_level.txt中

## 代码仓库地址

<https://github.com/zwx-zwx/sklearn-demo>