

# 法律声明

---

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象学院

■ 新浪微博：小象AI学院



# 催收评分卡模型的开发 -还款率模型介绍

# 目录

---

催收工作的简介

催收评分卡模型的简介

随机森林在还款率预测模型中的应用

# 催收工作的简介

## □ 信贷客户管理的周期



# 催收工作的简介

## □ 逾期客户的类型与风险等级

第一类：  
轻度

- 还款意愿和还款能力良好，客户因特殊原因（如生意出差在外地没赶回、忘了、特殊情况没时间等，出现数天逾期）

第二类：  
中轻度

- 还款意愿良好，还款能力出现问题

第三类：  
中度

- 无还款意愿或还款意愿恶化，有还款能力

第四类：  
重度

- 无还款意愿，还款能力弱化或完全丧失还款能力

# 催收工作的简介

---

## □ 催收流程(以信用卡为例)

第一步：短信催收。如果持卡人在还款日的7-15天内没有还款，银行会发短信提醒

第二步：电话催收。如果逾期不止一次，而且本次逾期超过了30天，那么银行会打电话

第三步：实地催收。如果逾期超过90天，银行就要上门要钱

第四步：诉诸法院。如果时间再长，银行通过法务提起诉讼

第五步：第三方催收。把逾期的资产打包卖给催收公司

# 目录

---

催收工作的简介

催收评分卡模型的简介

随机森林在还款率预测模型中的应用

# 催收评分卡模型的简介

---

## □ 催收评分卡的简介



还款率模型

账龄滚动模型

失联预测模型



# 催收评分卡模型的简介

---

## □ 失联预测模型

在逾期阶段，对于尚能联系到的人群预测其未来失联的概率

### ➤ 模型常用指标

- 逾期天数(day past due, DPD)
- 逾期金额占比(outstanding/limit)
- 个人信息(性别，年龄，收入，工作等等)
- 联系人关系(是否是夫妻、子女、同事、朋友)
- 运营商信息(在网时长、高频联系人)

# 催收评分卡模型的简介

---

## □ 账龄滚动模型

预测逾期人群从轻度逾期变成重度逾期的概率

### ➤ 模型常用指标

- 逾期天数(day past due, DPD)
- 历史还款率信息
- 个人信息(性别, 年龄, 收入, 工作等等)
- DBR(debt burden ratio)

# 催收评分卡模型的简介

---

## □ 还款率预测模型

预测经催收后，最终催收回的欠款的比率

### ➤ 模型常用指标

- 逾期天数(day past due, DPD)
- 历史还款率信息
- 个人信息(性别，年龄，收入，工作等等)
- DBR(debt burden ratio)
- 联系人关系(是否是夫妻、子女、同事、朋友)

# 目录

---

催收工作的简介

催收评分卡模型的简介

随机森林在还款率预测模型中的应用

# 随机森林在还款率预测模型中的应用

---

## □ 随机森林(Random Forest, RF)模型简介

集成模型的一种。元分类器是分类树或回归树，组合方式是bagging。

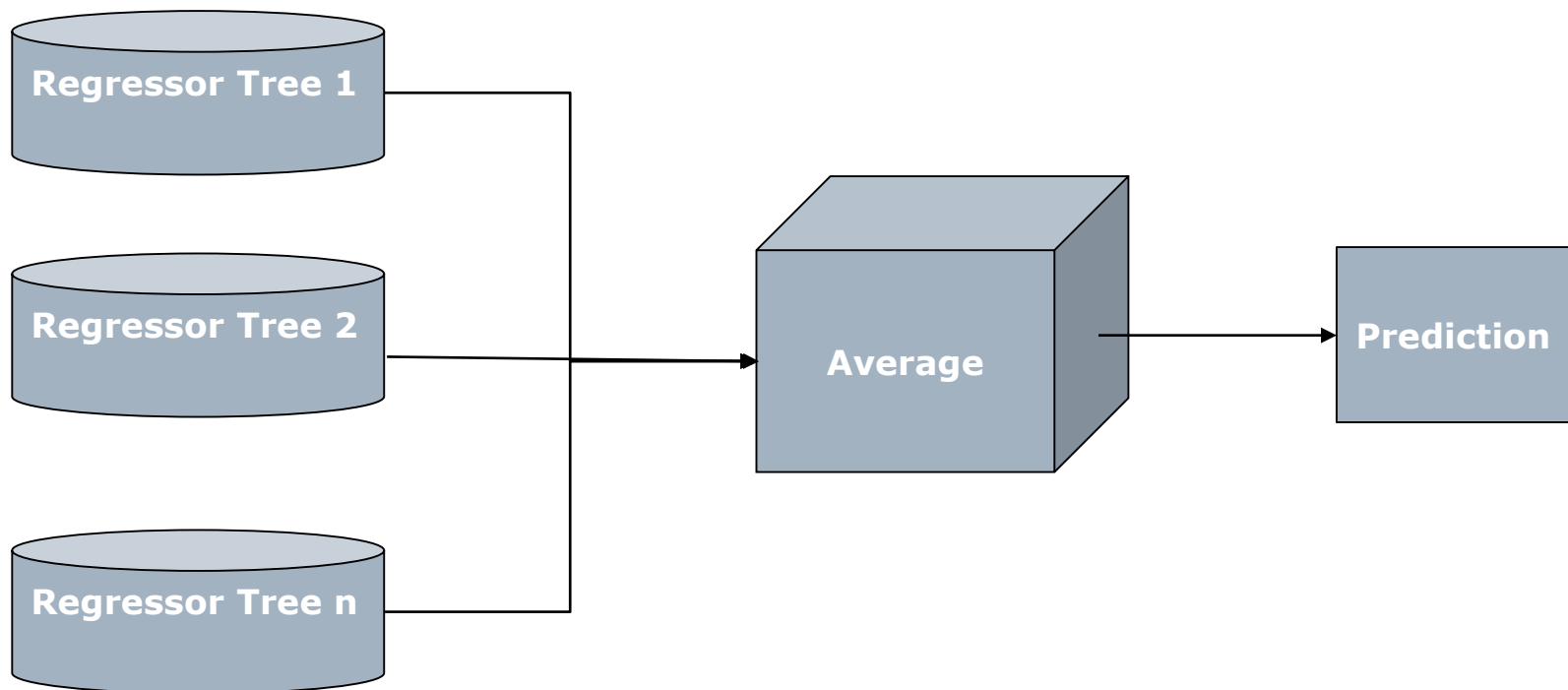
### ➤ 优点

- 泛化能力强
- 对数据质量要求不高
- 预测精度相对较高
- 可以评价变量的重要性
- 计算量不大，且可以并行计算

# 随机森林在还款率预测模型中的应用

## □ 随机森林(Random Forest, RF)模型简介

### 基于回归树的RF



# 随机森林在还款率预测模型中的应用

---

## □ 随机森林模型的训练步骤

基于回归树的RF:

- ① 从样本集中通过重采样的方式产生 $n$ 个样本
- ② 假设样本特征数目为 $a$ ，对 $n$ 个样本选择 $a$ 中的 $k$ 个特征，用建立回归树的方式获得最佳分割点
- ③ 重复 $m$ 次，产生 $m$ 棵回归树
- ④ 预测结果的平均值机制来进行预测

# 随机森林在还款率预测模型中的应用

---

## □ 案例

在Prosper 公司发生逾期的一批客群，数据包括

- 催收数据
- 逾期数据
- 申请阶段的信用评分
- 财产、收入类数据
- 其他信用类数据



# 随机森林在还款率预测模型中的应用

---

## □ 基于随机森林的还款率模型开发的步骤

① 数据清洗

② 基于回归树的RF模型调参

③ 模型测试

# 随机森林在还款率预测模型中的应用

## □ 步骤一：数据预处理

### ➤ 定义还款率：

$$\text{recovery rate} = \frac{\text{recovery amount}}{\text{delinquent amount} + \text{collection fee}}$$

### ➤ 数据缺失值处理：

- 对于类别型变量，视缺失为单独的值
- 对于联系型变量，用非缺失的均值进行补缺

### ➤ 类别型变量编码：

- 用同一类别的平均还款率作为编码的数值

# 随机森林在还款率预测模型中的应用

---

## □ RF模型的调参

RR模型最重要的参数

- 树的个数
- 树的最大深度
- 树的内部节点划分的最小样本数
- 叶子节点最少样本数
- 最大特征数

调参方法：网格搜索法

调参依据：均方误差

# 随机森林在还款率预测模型中的应用

---

## □ 预测结果

### ➤ 前5个最重要的变量：

BankcardUtilization': 0.039835423518574609,  
'BorrowerAPR': 0.031292686072593459,  
'BorrowerRate': 0.028968943224986696,  
'BorrowerState\_encoded': 0.059830669350693005,  
'CreditGrade\_encoded'

### ➤ 平均误差率：14%

# 疑问

---

## □ 小象问答官网

■ <http://wenda.chinahadoop.cn>

# 联系我们

---

## 小象学院：互联网新技术在线教育领航者

- 微信公众号：小象学院
- 新浪微博：小象AI学院

