

## 07 | 互联网金融：芝麻信用分的建模过程是怎样的？

### 1. 背景

- 1.1. 两种角色
  - 数据建模师
  - 风控分析师
- 1.2. 两种产品
  - to B：企业信用评估
  - to C：个人信用评分
- 1.3. 工作内容
  - 数据源
  - 信用评分模型
  - 模型监控维护
  - 其他数据挖掘



(这里以“芝麻信用分”为例进行解读)

### 2. 授信模型

- 2.1. 数据源
  - 在选择变量的时候，基于 RFM 原则，即最近、频次、钱。
- 2.2. 数据处理
  - 数值型和字符串型字段，缺失值和异常值检验
  - 数值型进行相关性验证，去掉强相关性字段
  - 字符串型字段做离散化处理，比如专家打分法
- 2.3. 数据标准化
  - Min-Max标准化
  - Z-score标准化
  - .....
- 2.4. 数据建模
  - 建模前，理清业务目标
  - 根据目标找到需要的模型
  - 抽样评估（这里使用混淆矩阵和ROC曲线）模型的离线效果
  - 模型运行周期，初期基本每月一次

### 3. 模型落地

- 落地前
  - 找落地场景，这里还是以“芝麻分申请招联金融信用额度”为例讲解：
  - 招联会给一批用户样本来让芝麻算这批用户的违约率
  - 招联根据违约率来算模型准确度
  - 准确度90%以上，双方才合作！
- 落地中
  - 正式落地时，招联会参考芝麻信用分、用户在招联信用分、央行信用评估值三方进行联合，出一套授信方案！
- 落地后
  - 招联会每周和芝麻对一次坏账！进行调参。
  - 模型最终的评估指标是坏账率！
  - 调参方法：
    - 确认是子模型还是整个模型问题；
    - 若是子模型，则直接调子模型参数；否则重新数据处理。