## 电销客户名单筛选总结

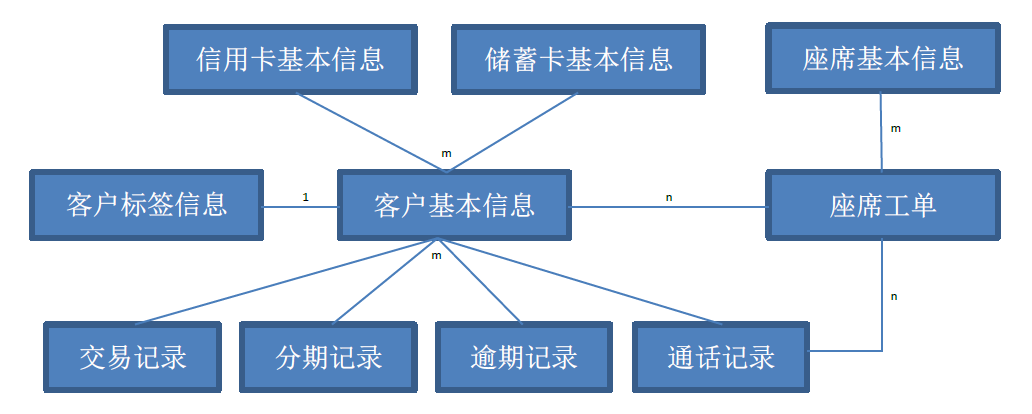
1. 产品简介

基于银行历史数据，利用数学模型，针对不同业务的电话销售客户名单进行筛选，目的是提高营销的精确度，降低电话销售成本，增加销售成功率。

1. 数据源

数据库中共存储12张数据表，如下所示：

根据数据表信息构建如下ER图：



1. 建模流程

（1）数据输入

（2）数据预处理

·原始数据转化（利用ETL工具）

·空缺值处理（RandomForest算法；平均值算法）

·变量均衡处理（SMOTE算法）

（3）模型算法原理

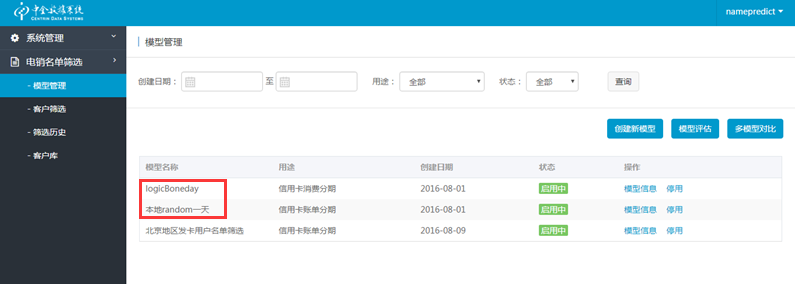
·随机森林算法 rf\_model.R；rf\_predict.R

·逻辑回归算法（算法筛选变量）glm\_calculate\_model.R；glm\_calculate\_predict.R

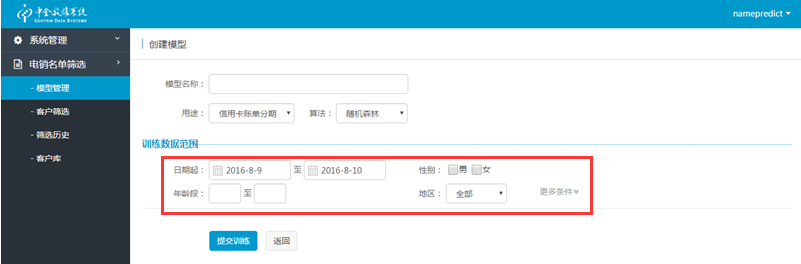
·逻辑回归算法（人工定义变量）glm\_input\_model.R；glm\_input\_predict.R

（4）模型训练模式

·可人工选择模型（随机森林、逻辑回归、组合算法结果）



·可人工选择数据范围（自定义时间、自定义性别、自定义地点等）

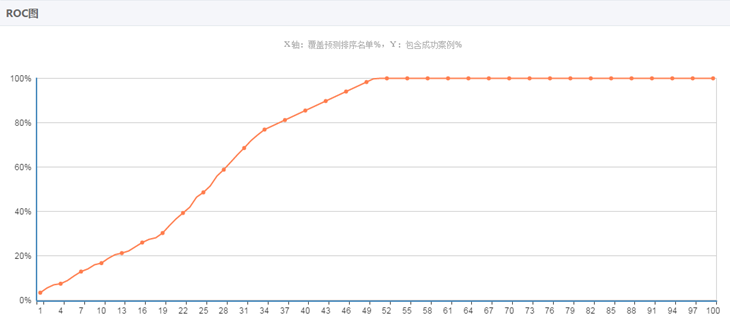


（5）模型评估

·成功情况统计（筛选客户量，成功客户量，成功率，AUC值）

·混淆矩阵（预测结果VS实际结果）

·ROC曲线（查看最佳电销名单筛选点）



·日走势图（模型在选定范围内，每日成功概率走势图）

1. 系统功能

包括4个方面：模型管理、客户筛选、筛选历史查询、客户列表

1. POC实施流程

|  |  |  |  |
| --- | --- | --- | --- |
| **流程步骤** | **工作事项** | **要求** | **负责方** |
| 环境准备 | 1、设备场地准备 2、网络环境准备 3、办公场地准备 | 服务器：1-2台,内存:32G+,cpu:6core,1200MHz FTP服务及机器:linux，windows ETL工具：kettle 算法库：3.2.1以上语言包 数据库：mysql5.1以上 （特别版本要求） | 中金、客户 |
| 系统部署 | 1、电销系统 | Jdk7,tomcat7,r语言 | 中金 |
| 客户测试数据引入 | 1、客户提供数据 2、数据初步分析 3、确定测试数据范围 | 1、数据量要达到几万级别 2、尽量减少缺失值，使数据完善 | 中金、客户 |
| 数据预处理 | 按照一定的规则对数据进行预处理，使数据符合一定的规范 | ETL工具 | 中金 |
| 模型训练及验证 | 1、筛选3万条的电销历史数据导入系统训练模型 2、查看模型效果 3、分三批验证模型效果，每批导入M个数据 4、分析效果 5、根据效果及训练数据优化模型 6、模型试用结果报告 | R语言 | 中金 |
| 投产试用 | 1、每天按照名单需求的20倍数输入客户数据进行预测 2、试用周期为五个工作日，其中第三日可根据模型前两日的表现对模型进行调整 |  | 中金、客户 |
| 试用效果评估 | 1、试用报告 2、根据试用过程中模型的表现，决定继续优化或直接投产使用 |  | 中金、客户 |
| 模型的优化 | 依据试用结果，继续优化系统 |  | 中金 |
| 正式投产 | 正式将系统投入使用 |  | 中金、客户 |
| 持续优化 | 在客户使用过程中，对系统进行维护与优化 |  | 中金、客户 |