

数据挖掘-快速入门

- 案例背景
 - 1. 新建实验
 - 2. 选择数据源
 - 3. 模型搭建
 - 4. 训练&预测
 - 5. 结果评估

Smartbi Mining为用户提供拖拽的流式建模、可视化建模界面、预置大量数据处理及算法节点，方便用户快速构建数据挖掘模型，通过简单的例子来看一下如何使用Smartbi Mining搭建实验。[点击入门视频下载](#)

本文档将按照如下流程，以客户流失问题为背景，为您介绍如何快速在Smartbi Mining上搭建实验：



案例背景

现在银行产品同众化现象普遍存在，客户选择产品和服务的途径越来越多，客户对产品的忠诚度越来越低，所以客户流失已经成为银行业最关注的问题之一。其实，银行流水数据反应了客户使用银行卡的行为特点，通过对这些数据进行深入分析，挖掘不同类别客户的特点，有针对性地进行营销，加强客户关系管理，提高客户与我行粘度，减少不必要的客户流失。

本文档主要介绍利用Smartbi Mining快速搭建挖掘实验的过程，故对原始数据进行一定分析之后，加工必要数据后，选取影响较大的字段数据作为本文档构建模型的数据输入，故本实验构建过程无数据预处理过程，相关详情参考 [内置案例-银行客户流失](#)。

现有银行客户流失数据84823条，每条数据记录包括客户id、卡等级、是否为代发客户、月均代发金额、最多代发金额、月初AUM、月均AUM、是否流失8个字段，各字段解释如下：

字段名	数据类型	字段解释
客户id	string	客户唯一身份标识
卡等级	int	客户拥有卡的级别，总共分为[1, 2, 3, 4]四个级别
是否为代发客户	int	1表示为代发客户，0表示为非代发客户
月均代发金额	int	代发客户的月代发金额平均值
最多代发金额	int	代发客户代发金额的最大值
月初AUM	int	客户月初的资产管理规模AUM值
月均AUM	int	客户资产管理规模AUM的月均值
是否流失	int	1表示该客户流失，0表示该客户未流失。

1. 新建实验

数据挖掘的默认界面为‘实验管理’界面，该界面可以进行‘新建实验’：

新建方式1：在文件夹右键菜单‘新建’中单击‘实验’；

新建方式2：单击‘新建实验’。



2. 选择数据源

根据原始数据存放的方式在选择数据源节点，将数据源节点拖拽至画布区域，并在参数面板上进行数据源信息的设置。

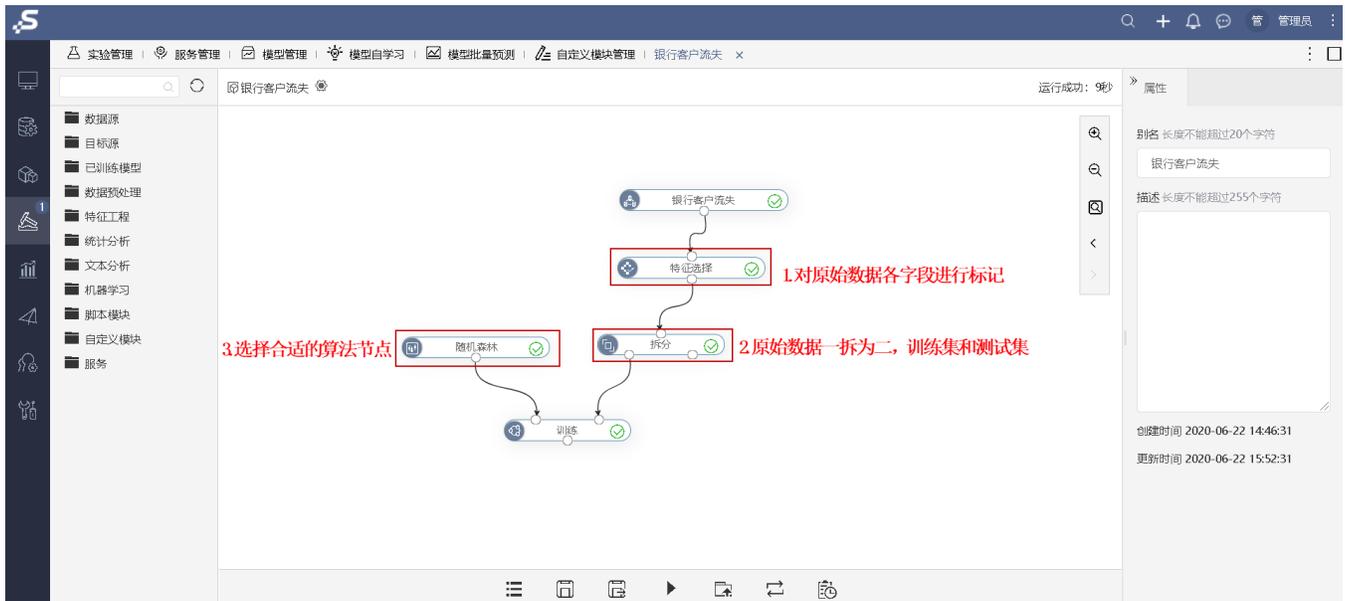


3. 模型搭建

特征选择: 将‘卡等级’、‘是否为代发客户’、‘月均代发金额’、‘最多代发金额’、‘月初AUM’、‘月均AUM’做为特征列，而‘是否流失’做为标签列；

拆分: 将数据集进行拆分，一部分数据用来训练模型，另一部分用来验证经过训练的模型效果如何；

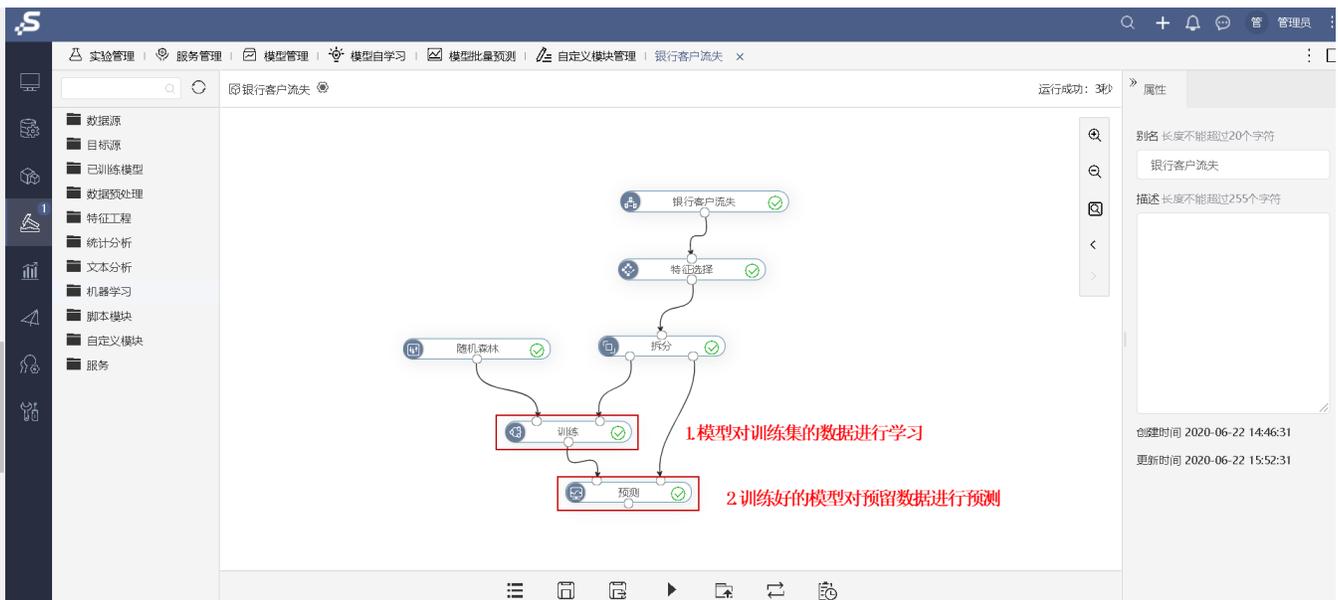
随机森林: 本研究用分类算法节点中选择 [随机森林算法](#) 进行预测客户流失的模型构建。



4. 训练&预测

训练: 训练节点的输入分为两部分，左边为算法节点，右边为训练数据集，模型通过对训练集数据中特征列的数据进行学习，并根据标签列的数据，确认具有何种特征的数据可能是流失客户，进而使模型获得能够识别流失客户的能力；

预测: 预测节点的输入分为两部分，左边为训练好的模型，右边为测试数据集，模型运用识别流失客户的能力，对测试集中的特征列数据进行处理，并获得是否为流失客户的标记，这里的标记是模型给出的结论，并非测试集中标签列的数据。



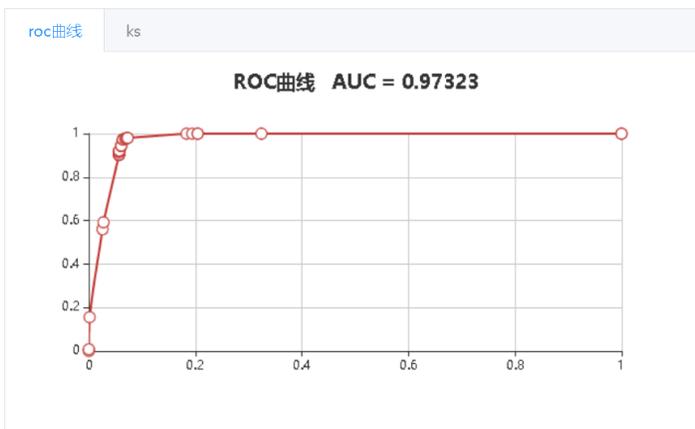
5. 结果评估

评估: 通过对比模型预测结果与测试集数据之间的差异，可以评价模型的效果。

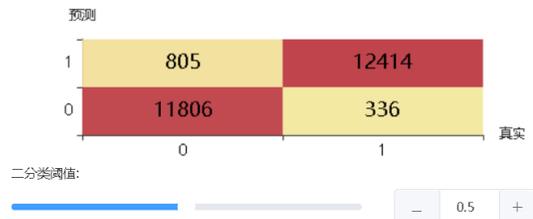
通过比较预测值和真实值之间的差异, 对模型效果给出评价

最终输出评价结果为:

[查看分析结果](#)



指标	值
ACC	0.95501
precision	0.97233
F1 score	0.9539
recall	0.93617
conf	0.92625



可以看到这个模型的准确率高达95%。