

产品实施流程

Smartbi Mining凝聚丰富的数据挖掘项目经验，也提炼出科学的实施流程。首先是确定商业目标，这个目标同时要具备业务定义与技术定义（例如我们前面提到的，预测商业银行客户的流失概率。就是一个明确的目标）；第二步是 进行数据的准备与模型建立。（这两者其实是统一的，使用什么算法建立模型其实与具备什么业务数据密切相关。这里面会进行不断的实验与调整）接下来，模型相对稳定后，就需要用不同的评估方法进行模型的评估与参数调优。在达到一定的准确率后（比如8成），模型就能进行部署，如前面所述这个在我们产品里面就很简便了。而部署并不是终结，挖掘项目通常都需要迭代。如 调整目标、补充数据、 优化模型性能等，这也是我们很多项目都是多期的。

1) 第一步确定商业目标：

基于业务的现状，我们要明确我们的分析目标是什么？从需求现状出发，挖掘出潜在的需求。

比如在银行精准营销案例中，

业务现状：银行推荐某个理财产品的转换率低

需要解决的问题：提高理财产品的转换率

最终希望得到的结果：转换率达到xx%

2) 准备数据：

准备数据包括了数据采集，数据预处理等一系列动作，是为了数据建模做准备，通常是很耗时耗力的工作，需要足够的耐心。

数据取样：首先我们根据我们的商业目标，收集相关的、有可能影响的数据。

采集数据时候需要充分考虑以下四个问题：

- 指标：如何设计指标体系？
- 数据：需要什么数据？
- 样本量：分析的样本量多少合适？
- 取数周期：使用哪个阶段的数据

数据处理：数据处理不仅仅是删除错误数据或插入缺失值，还包括查找数据中的隐含相关性、标识最准确的数据源并确定哪些列最适合用于分析。通常的流程包含数据探索、数据修正和特征变换，并且是反复循环的进行。

3) 建立模型：

建立模型是数据挖掘工作的核算环节，建立模型是对采样数据的轨迹的概况，具体表现就是预测的公式，公式产出与观察值有相似结构的输出就是预测值。构建模型的前提是在样本数据中发现模式，比如本次建模是属于挖掘应用中的哪类问题（分类、回归、聚类、关联规则），确定了问题后选用那种算法进行模型的构建？必要时可以同时采用集中预测模型进行运算以便对比、选择。

建立模型通常包含：模型的建立、训练、验证、预测几个步骤。

4) 模型评价：

建立了模型就要检测模型好不好，模型评价的目的是从构建的模型里面找出最好的一个模型， 模型评价从两方面考虑，首先从模型本身参数评估，包括准确率、查准率、召回率、F1得分、R2。另外还要从业务角度去评估。

5) 应用部署：

挖掘项目中，将模型部署生产是最后一公里，但这一般是一个繁琐的过程，但是在Smartbi Mining中这变得非常轻松。部署后，其他的应用就能通过接口来调用，获得预测值。最后管理员还能通过界面监控这个接口的情况。