

大数据分析挖掘综合能力提升实战

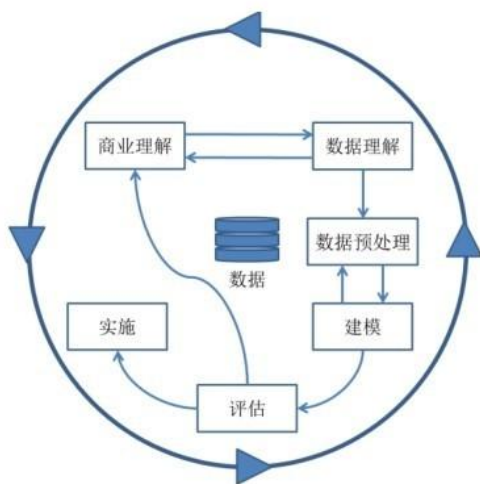
【课程目标】

本课程为高阶课程，面向所有业务支撑部门及数据分析部门。

本课程的主要目的是，帮助学员掌握一些业务专题挖掘模型，帮助学员建立对复杂业务问题的数据挖掘综合能力。

本课程具体内容包括：

- 1、数据挖掘流程，数据预处理
- 2、用户专题分析：用户群划分/客户价值评估/客户偏好分析/用户行为预测
- 3、产品专题分析：产品设计优化、产品功能评估、产品最优定价策略
- 4、精准推荐算法：协同过滤、关联分析、基于内容/用户的推荐(CBR/UBR)
- 5、金融风险评估：信用评分卡模型、风险预测模型



- 1 市场客户细分
- 2 客户价值评估
- 3 客户喜好评估
- 4 产品设计分析
- 5 产品定价策略
- 6 精准推荐算法
- 7 购物篮关联分析
- 8 信用评分卡模型

本系列课程从实际的业务需求出发，结合行业的典型应用特点，围绕实际的商业问题，对数据预测建模的过程进行了全面的介绍（从模型选择，到特征选择，再到训练模型，评估模型，以及优化模型和模型解读），通过大量的操作演练，帮助学员掌握数据建模的思路、方法、技巧，以提升学员的数据建模的能力，支撑运营决策的目的。

通过本课程的学习，达到如下目的：

- 1、熟悉数据挖掘的标准过程，熟悉每个步骤的具体操作。
- 2、掌握数据预处理的任务，熟练使用 SPSS 工具完成预处理。
- 3、熟练掌握常用的业务专题分析模型：
 - a) 学会做市场客户细分，划分客户群
 - b) 学会实现客户价值评估

- c) 学会产品功能设计与新产品销量预测
- d) 熟悉产品定价策略，寻找产品最优定价
- e) 熟悉精准推荐策略，学会精准推荐产品
- f) 掌握信用评分卡的模型构建

【授课时间】

2-3天时间（每天6个小时）

【授课对象】

业务支撑部、运营分析部、数据分析部、大数据系统开发部等对业务数据分析有较高要求的相关人员。

【学员要求】

- 1、 每个学员自备一台便携机(必须)。
- 2、 便携机中事先安装好 Microsoft Office Excel 2013 版本及以上。
- 3、 便携机中事先安装好 IBM SPSS Statistics v24 版本及以上。

注：讲师可以提供试用版本软件及分析数据源。

【授课方式】

数据分析基础 + 方法讲解 + 实际业务问题分析 + 工具实践操作

采用互动式教学，围绕业务问题，展开数据分析过程，全过程演练操作，

让学员在分析、分享、讲授、总结、自我实践过程中获得能力提升。

【课程大纲】

第一部分：数据挖掘流程—挖掘步骤篇

- 1、 数据挖掘概述
- 2、 数据挖掘的标准流程（CRISP-DM）
 - 商业理解
 - 数据准备
 - 数据理解
 - 模型建立
 - 模型评估
 - 模型应用

案例：客户流失预测及客户挽留

- 3、 数据集概述

- 4、 SPSS 工具介绍
- 5、 数据挖掘常用模型

第二部分：数据挖掘流程—数据预处理

如何整理数据，了解数据，对数据进行预处理？

- 1、 数据预处理的四大任务
 - 数据集成：多个数据集合并
 - 数据清洗：异常值的处理
 - 样本处理：样本筛选、样本抽样、样本平衡
 - 变量处理：变量变换、变量派生、变量精简
- 2、 数据集成（数据集合并）
 - 样本追加（添加数据行）：横向合并
 - 变量合并（添加变量列）：纵向合并
- 3、 数据清洗（异常数据处理）
 - 取值范围限定
 - 重复值处理
 - 无效值/错误值处理

- 缺失值处理
- 离群值/极端值处理
- 数据质量评估

4、 样本处理：行处理

- 样本筛选：指定条件筛选指定样本集（减少样本数量）
- 样本抽样：随机抽取部分样本集（减少样本数量）
- 样本平衡：正反样本比例均衡

5、 变量处理：列处理

- 变量变换：原变量取值更新，比如标准化
- 变量派生：根据旧变量生成新的变量
- 变量精简：变量删除/降维，减少变量个数
- 类型转换：数据类型的相互转换

6、 变量精简/变量降维常用方法

- 常用降维方法
- 如何确定降维后变量个数
- 特征选择：选择重要变量，剔除不重要变量
 - ◇ 基于变量本身特征来选择属性

- ◇ 基于数据间的相关性来选择属性
- ◇ 利用 IV 值筛选
- ◇ 基于信息增益来选择属性
- 因子合并：将多个变量进行合并
 - ◇ PCA 主成分分析
 - ◇ 判别分析

7、 类型转换

8、 因子合并/主成分分析

- 因子分析的原因
- 因子个数选择原则
- 如何解读因子含义

案例：提取影响电信客户流失的主成分分析

9、 数据探索性分析

演练：描述性分析（频数、描述、探索、分类汇总）

10、 数据可视化

演练：各种图形绘制

第三部分：市场细分模型—聚类模型篇

问题：我们的客户有几类？各类特征是什么？如何实现客户细分，开发符合细分市场的新产品？如何提取客户特征，从而对产品进行市场定位？

1、市场细分的常用方法

- 有指导细分
- 无指导细分

2、聚类分析

- 如何更好的了解客户群体和市场细分？
- 如何识别客户群体特征？
- 如何确定客户要分成多少适当的类别？
- 聚类方法原理介绍
- 聚类方法作用及其适用场景
- 聚类分析的种类
 - ◇ K均值聚类
 - ◇ 层次聚类
 - ◇ 两步聚类
- K均值聚类（快速聚类）

案例：移动三大品牌细分市场合适吗？

演练：宝洁公司如何选择新产品试销区域？

演练：如何自动评选优秀员工？

演练：中国各省份发达程度分析，让数据自动聚类

- 层次聚类（系统聚类）：发现多个类别
- R型聚类与Q型聚类的区别

案例：中移动如何实现客户细分及营销策略

演练：中国省市经济发展情况分析（Q型聚类）

演练：裁判评分的标准衡量，避免“黑哨”（R型聚类）

- 两步聚类

3、客户细分与PCA分析法

- PCA主成分分析的原理
- PCA分析法的适用场景

演练：利用PCA对汽车客户群进行细分

演练：如何针对汽车客户群设计汽车

第四部分：客户价值评估—RFM 模型篇

营销问题：如何评估客户的价值？不同的价值客户有何区别对待？

1、如何评价客户生命周期的价值

- 贴现率与留存率
- 评估客户的真实价值
- 使用双向表衡量属性敏感度
- 变化的边际利润

案例：评估营销行为的合理性

2、RFM 模型（客户价值评估）

- RFM 模型，更深入了解你的客户价值
- RFM 模型与市场策略
- RFM 模型与活跃度分析

演练：“双 11”淘宝商家如何选择价值客户进行促销

演练：结合响应模型，宜家 IKEA 实现最大化营销利润

案例：重购用户特征分析

第五部分：产品设计优化—随机效用理论

1、产品专题分析主要任务

- 产品设计分析
- 市场占有分析
- 累计销量分析
- 定价策略分析

2、产品设计优化（联合分析法）

问题：如何设计最优的功能特征？

- 评估功能特征的重要性
- 评估功能特征的价值

案例：产品开发与设计分析

3、产品评估模型（随机效用理论）

- 属性重要性评估
- 市场占有率评估
- 产品价格弹性评估
- 评估产品的品牌价值
- 动态调价（纳会均衡价格）

案例：品牌价值与价格敏感度分析

案例：纳什均衡价格

第六部分：产品定价策略—最优定价篇

营销问题：产品如何实现最优定价？套餐价格如何确定？采用哪种定价策

略可达到利润最大化？

1、常见的定价方法

2、产品定价的理论依据

- 需求曲线与利润最大化
- 如何求解最优定价

案例：产品最优定价求解

3、如何评估需求曲线

- 价格弹性
- 曲线方程（线性、乘幂）

4、如何做产品组合定价

5、如何做产品捆绑/套餐定价

- 最大收益定价（演进规划求解）

- 避免价格反转的套餐定价

案例：电信公司的宽带、IPTV、移动电话套餐定价

6、非线性定价原理

- 要理解支付意愿曲线
- 支付意愿曲线与需求曲线的异同

案例：双重收费如何定价（如会费+按次计费）

7、阶梯定价策略

案例：电力公司如何做阶梯定价

8、数量折扣定价策略

案例：如何通过折扣来实现薄利多销

9、定价策略的评估与选择

案例：零售公司如何选择最优定价策略

10、 航空公司的收益管理

- 收益管理介绍
- 如何确定机票预订限制
- 如何确定机票超售数量
- 如何评估模型的收益

案例：FBN 航空公司如何实现收益管理（预订/超售）

第七部分：产品推荐算法—推荐模型篇

问题：购买 A 产品的顾客还常常要购买其他什么产品？应该给客户推荐什么产品最有可能被接受？

1、从搜索引擎到推荐引擎

2、常用产品推荐模型及算法

3、基于流行度的推荐

- 基于排行榜的推荐，适用于刚注册的用户
- 优化思路：分群推荐

4、基于内容的推荐 CBR

- 关键问题：如何计算物品的相似度
- 优缺点
- 优化：Rocchio 算法、基于标签的推荐、基于兴趣度的推荐

5、基于用户的推荐

- 关键问题：如何对用户分类/计算用户的相似度
- 算法：按属性分类、RFM 模型、PCA、聚类、按偏好分类、按地理位置

6、协同过滤的推荐

- 基于用户的协同过滤
- 基于物品的协同过滤
- 冷启动的问题

案例：计算用户相似度、计算物品相似度

7、基于关联分析的推荐

- 如何制定套餐，实现交叉/捆绑销售

案例：啤酒与尿布、飓风与蛋挞

- 关联分析模型原理 (Association)
- 关联规则的两个关键参数
 - ◇ 支持度
 - ◇ 置信度
- 关联分析的适用场景

案例：购物篮分析与产品捆绑销售/布局优化

案例：通信产品的交叉销售与产品推荐

8、基于分类模型的推荐

9、其它推荐算法

- LFM 基于隐语义模型

- 按社交关系

- 基于时间上下文

10、 多推荐引擎的协同工作

第八部分： 银行信用卡评估—信用评分卡模型

1、 信用评分卡模型简介

2、 评分卡的关键问题

3、 信用评分卡建立过程

- 筛选重要属性

- 数据集转化

- 建立分类模型

- 计算属性分值

- 确定审批阈值

4、 筛选重要属性

- 属性分段

- 基本概念：WOE、IV

- 属性重要性评估

5、数据集转化

- 连续属性最优分段
- 计算属性取值的 WOE

6、建立分类模型

- 训练逻辑回归模型
- 评估模型
- 得到字段系数

7、计算属性分值

- 计算补偿与刻度值
- 计算各字段得分
- 生成评分卡

8、确定审批阈值

- 画 K-S 曲线
- 计算 K-S 值
- 获取最优阈值

案例：构建银行小额贷款的用户信用模型

结束：课程总结与问题答疑。