

# 大数据挖掘与分析培训纲要

<b>内容概要</b>	1.介绍 SPSS Modeler 数据挖掘软件的功能和基本操作； 2.学习如何使用 SPSS Modeler 完成数据挖掘过程； 3.学习机器学习的常见技术，例如神经网络、决策树、聚类、关联分析、回归分析和社会网络用户情感分析等，并学习如何将它们应用到业务数据。 4.结合实际案例以及上机操作讲解
<b>培训内容</b>	1. 数据挖掘简介 数据挖掘的概念 CRISP-DM 方法论 SPSS Modeler 18 简介 2. 数据理解 数据理解 缺失值定义 数据审核节点介绍 分布图节点----初步理解字符型字段的分布 直方图/统计节点----初步理解数值型字段的分布 常用数据分析图画法 3. SPSS Modeler 分类技术 决策树技术 Logistics 回归 神经网络 贝叶斯分类器 4. SPSS Modeler 细分技术 Kohonen 网络/两步聚类/K-means 5. SPSS Modeler 关联分析技术 常见的回归分析 Carma/Apriori 序列节点 6. SPSS Modeler 自动建模技术 自动聚类节点 自动分类节点 自动数值节点 7. SPSS Modeler 其他技术 线性回归 特征选择 支持向量机 (SVM) 主成分/因子分析 8. SPSS Modeler 模型评估技术 评估图 分析节点

	<p>案例 1：银行客户数据分析</p>	<p>以客户信用记录、消费历史记录数据为基础，使用 IBM SPSS Modeler，分析决定用户信用等级的主要因素是什么？并给与解释。信用额度和信用评分有何关系？给出欺诈的判断模型。分析客户人口属性对银行信贷和信用评分有无影响。分析逾期还款的客户有什么特征？根据消费历史分析各类客户的特征。</p>
	<p>案例 2：大众点评网的用户偏好分析</p>	<p>到大众点评网抓取多家火锅店的店名、位置、点评条数、人均消费额、口味评分（均分）、环境评分（均分）、服务评分（均分）、点评的均分、有无团购、有无送外卖、有无订座等数据（还可以采集标签等数据），对火锅店进行分类、找出每类火锅店的特征，并分析点评均分与其他变量之间的关系。为新开设火锅店选址提供指导；为用户推荐火锅店。</p>
	<p>案例 3：制造企业质量数据挖掘</p>	<p>本项目在中天耐热导线工厂最近 2 年的质量管理数据的基础上，分析了这些数据存在的问题，进行了大量的预处理，利用统计学、多维度分析、数据挖掘以及可视化等多种数据分析方法，以优化耐热导线的制造过程。</p>
	<p>案例 4：电商客户情感分析</p>	<p>通过抓取对典型电商客户的评论，分析客户对于某个商品多个维度的态度和情感倾向，以及客户注重商品属性所在，商品的卖点、不足以及同类商品竞争力对比。</p>