

# 《SPSS Modeler 大数据挖掘方法与应 用实战操作》培训大纲

## 壹、 课程内容与安排：

模块一：概念篇—数据挖掘的概念与流程及方法

模块二：操作篇—SPSS Modeler 数据挖掘工具操作方法及运用

模块三：算法与建模篇—SPSS Modeler 数据挖掘算法及数学建模方法

模块四：实战演练篇—SPSS Modeler 算法及建模实战演练与操作

模块五：高级进阶篇—运用 SPSS Modeler 关联分析法提升 4G 网络满意度

模块六：撰写分析报告篇—如何撰写漂亮的数据分析报告

## 贰、 课程时间：3 天，6 小时/天

参、 授课对象：经分分析相关人员、客户响应业务支撑人员、数据和 IT 等部门的管理人员及技术人员等。

## 四、 授课方式：

理论讲授，案例分析，方法传授、动画演示、互动讨论，讲师点评、实战演练。

## 伍、 课程纲要：

模块一：概念篇—数据挖掘的概念与流程及方法

### 1. 数据挖掘的概念及目标

1.1 数据挖掘的定义及功能

1.2 数据挖掘的基本特征

1.3 数据挖掘的知识表示

1.4 数据挖掘的目标

1.5 数据挖掘的主要任务

1.6 数据挖掘与大数据关系

- 1.7 数据挖掘在电信行业的应用内容
- 2. 数据挖掘与相关领域之间的关系
  - 2.1 数据挖掘与专家系统的关系
  - 2.2 数据挖掘与统计分析的关系
  - 2.3 数据挖掘与人工智能的关系
  - 2.4 数据挖掘与机器学习
  - 2.5 数据挖掘与数据仓库
  - 2.6 数据挖掘与 OLAP
  - 2.7 数据挖掘与统计学
  - 2.8 数据挖掘与智能决策
  - 2.9 数据挖掘与云计算
- 3. 数据挖掘的对象-在何种数据上进行数据挖掘
  - 3.1 关系数据库
  - 3.2 数据仓库
  - 3.3 事务数据库
  - 3.4 高级数据库系统
  - 3.5 展开文件和 WWW
- 4. 数据挖掘的分类
  - 4.1 根据挖掘的数据库类型分类(不同标准如数据模型、涉及应用类型)
  - 4.2 根据挖掘的知识类型分类 (不同功能如特征化、区分、关联等)
  - 4.3 根据所用的技术分类 (如机器学习、统计学、可视化、模式识别)
  - 4.4 根据应用分类 (如金融、电信、股票市场、DNA、e-mail 等)
- 5. 数据挖掘功能-可以挖掘什么类型的模式
  - 5.1 概念/类描述：特征化和区分 (Characterization and Discrimination)
  - 5.2 关联分析 (Association Analysis)
  - 5.3 分类和预测 (Classification and Predict)
  - 5.4 聚类分析 (Clustering Analysis)
  - 5.5 孤立点分析 (Outlier Analysis)
  - 5.6 演变分析 (Evolution Analysis)
- 6. 数据挖掘问题的描述
  - 6.1 描述需要解决的关键问题
  - 6.2 描述如何转换成为数据挖掘

- 6.3描述数据挖掘算法的选择依据
- 6.4描述预测类还是描述类
- 6.5描述与各种算法的使用特点结合
- 7. **CRISP-DM 数据分析挖掘方法论介绍**
  - 7.1描述商业理解——要实现什么“目的”？
  - 7.2描述数据的理解以及收集——手里有哪些数据？
  - 7.3描述数据的准备——数据的清洗及转换
  - 7.4描述应用数据挖掘工具建立模型——使用哪种数据挖掘算法和工具？
  - 7.5描述模型评估——算法评估
  - 7.6描述部署（并形成数据挖掘报告）——实际使用及形成报告
- 8. **数据挖掘常用算法及案例分析**
  - 8.1分类算法及案例分析
  - 8.2预测型方法及案例分析
  - 8.3关联分析法及案例分析
  - 8.4聚类分析法及案例分析
  - 8.5回归分析法及案例分析
  - 8.6时序分析法及案例分析
  - 8.7智能推荐分析法及案例分析
  - 8.8神经网络分析法及案例分析
- 9. **数据挖掘建模过程**
  - 9.1数据探索
  - 9.2数据预处理
  - 9.3挖掘建模
  - 9.4模型评估
- 10. **数据挖掘流程**
  - 10.1 数据挖掘思路
  - 10.2 数据挖掘项目立项
  - 10.3 数据挖掘项目实施
    - 10.3.1 业务理解阶段 (BUSINESS UNDERSTANDING)
    - 10.3.2 数据理解阶段 (DATA UNDERSTANDING)
    - 10.3.3 数据准备阶段 (DATA PREPARATION)
    - 10.3.4 建模阶段 (MODELING)

**10.3.5 模型评估阶段 (EVALUATION)**

**10.3.6 部署阶段 (DEPLOYMENT)**

**11.数据分析：由上至下梳理数据分析体系**

**11.1.1 明确目标**

**11.1.2 分解指标**

**11.1.3 细化字段**

**11.1.4 非功能要求**

**11.1.5 系统实施**

**12.成果输出：由下至上实施落地到应用系统**

**12.1.1 连接数据**

**12.1.2 数据处理**

**12.1.3 数据建模**

**12.1.4 制作数据报告**

**12.1.5 非功能需求实现**

**13.数据挖掘的应用领域及案例**

**13.1.1 “电信运营”数据挖掘应用**

**13.1.2 “互联网与电子商务行业”数据挖掘应用**

**13.1.3 “政府”数据挖掘应用**

**13.1.4 “金融业”数据挖掘应用**

**13.1.5 “零售业”数据挖掘应用**

**13.1.6 “教育业”数据挖掘应用**

**13.1.7 “医疗业”数据挖掘应用**

**13.1.8 “能源业”数据挖掘应用**

**13.1.9 “制造业”数据挖掘应用**

**13.1.10 “交通物流业”数据挖掘应用**

**模块二：操作篇—SPSS Modeler 数据挖掘工具操作方法及运用**

**1. 从一个完整案例操作入手：4G 目标客户分析**

**1.1需求分析 (需求动机调查、购买行为调查)**

**1.2研究方法 (根据客户信息、选择数据挖掘算法、建立数学模型)**

**1.3数据分析**

**1.3.1 录入数据**

- 1.3.2 定义数据源 (数据源设置、数据类型设置)
- 1.3.3 数据理解 (字段的理解及使用、对数据进行探索分析)
- 1.3.4 数据准备 (字段过滤、数据抽样设置、模型数据分流、预测字段设置、“特征选择”模型设置、模型预览等)
- 1.3.5 建立模型 (选择算法模型、添加算法组件、生成模型)
- 1.3.6 模型评估 (模型评估设置、模型分析和评估、模型增益评估、模型预测准确率及数值修正)
- 1.3.7 研究结论 (给出分析结果、进行目标分析)

## 2. SPSS Modeler 软件介绍

### 2.1 SPSS Modeler 的功能及特点

### 2.2 SPSS Modeler 的版本

### 2.3 SPSS Modeler 支持的算法

### 2.4 SPSS Modeler 的行业应用

#### 2.4.1 通信行业

#### 2.4.2 政府行业

#### 2.4.3 金融行业

#### 2.4.4 制造行业

#### 2.4.5 医药卫生

#### 2.4.6 教育科研

#### 2.4.7 市场调研

#### 2.4.8 应用模型

#### 2.4.9 连锁零售

## 3. SPSS Modeler 的安装与基础操作

### 3.1 SPSS Modeler 软件的安装

### 3.2 数据流基本操作

#### 3.2.1 生成数据流

#### 3.2.2 添加和删除节点

#### 3.2.3 连接数据流

#### 3.2.4 修改连接节点

#### 3.2.5 执行数据流

## 4. SPSS Modeler 数据挖掘方法及高级操作

### 4.1 SPSS Modeler 数据挖掘一般流程

#### 4.1.1 数据录入

#### 4.1.2 定义数据源



- 4.10 图形生成与制作**
  - 4.10.1 散点图**
  - 4.10.2 直方图**
  - 4.10.3 网络图**
  - 4.10.4 评估图**
- 4.11 模型评估**
  - 4.11.1 模型评估设置**
  - 4.11.2 模型分析和评估**
  - 4.11.3 模型增益评估**
  - 4.11.4 模型预测准确率**
  - 4.11.5 数值修正**
- 4.12 研究结论**
  - 4.12.1 给出分析结果**
  - 4.12.2 进行目标分析**
- 4.13 案例分析：4G 网络 KPI 指标分析**
  - 4.13.1 研究方法**
  - 4.13.2 数据分析**
  - 4.13.3 研究结论**

### **模块三：算法与建模篇—SPSS Modeler 数据挖掘算法及数学建模方法**

#### **1. SPSS Modeler 数据挖掘算法及数学建模方法介绍**

- 1.1 分类分析**
- 1.2 聚类分析**
- 1.3 关联分析**
- 1.4 线性回归分析**
- 1.5 时间序列分析**
- 1.6 因子分析**
- 1.7 决策树分析**
- 1.8 判别分析**
- 1.9 人工神经分析**
- 1.10 贝叶斯网络分析**
- 1.11 社交网络分析**

#### **2. 案例一：时间序列分析—PON 网络带宽利用率预测**

- 2.1 目标与要求**
  - 2.1.1 理解时间序列分析法的基本理论**
  - 2.1.2 掌握时间序列分析法的建模步骤**
  - 2.1.3 掌握使用 SPSS Modeler 进行时间序列分析的基本方法**
  - 2.1.4 学会使用时间序列模型节点进行相应分析**
- 2.2 研究方法：根据时间序列模型分析某地区未来三个月电信市场 PON 网络带宽利用率的预测值**

- 2.3 定义数据源
- 2.4 数据理解
- 2.5 数据准备
- 2.6 建立模型
  - 2.6.1 模型定义
  - 2.6.2 模型应用
- 2.7 建模步骤
  - 2.7.1 搜集历史资料并加以整理，编程时间序列，根据时间序列绘成统计图
  - 2.7.2 分析时间序列
  - 2.7.3 求时间序列的长期趋势、季节变动和不规则变动的值，并选定近似的数学模型来代表它们
  - 2.7.4 利用时间序列资料求出长期趋势、季节变动和不规则变动的数学模型
  - 2.7.5 根据模型进行预测、评估和分析
- 2.8 数据分析与模型评估
  - 2.8.1 模型残差
  - 2.8.2 添加散点图
  - 2.8.3 实际值与预测值比较
  - 2.8.4 模型预测值比较
- 2.9 研究结论
- 3. 案例二：决策树分析—电信客户流失分析
  - 3.1 目标与要求
    - 3.1.1 理解决策树分析法的基本理论
    - 3.1.2 掌握决策树分析法的建模步骤
    - 3.1.3 掌握使用 SPSS Modeler 进行决策树分析的基本方法
    - 3.1.4 学会使用 C5.0 节点进行决策树分析
  - 3.2 研究方法：运用决策树算法中的 C5.0 算法得到电信客户流失预测模型，根据电信客户资料，对数据进行挖掘，得出客户流失的原因和流失原因的大体分布。
  - 3.3 定义数据源
  - 3.4 数据理解
  - 3.5 数据准备
  - 3.6 建立模型
    - 3.6.1 模型定义
    - 3.6.2 模型应用
  - 3.7 建模步骤
    - 3.7.1 绘制树状图
    - 3.7.2 将各状态概率及损益值标于概率枝上
    - 3.7.3 计算各个方案期望值
    - 3.7.4 进行剪枝，比较各个方案的期望值
    - 3.7.5 根据模型进行预测、评估和分析
  - 3.8 数据分析与模型评估
    - 3.8.1 编译模型“分析”节点
    - 3.8.2 执行“分析”节点
  - 3.9 研究结论
- 4. 案例三：回归分析—电信客户流失因素分析

- 4.1 目标与要求
  - 4.1.1 理解回归分析法的基本理论
  - 4.1.2 掌握回归分析法的建模步骤
  - 4.1.3 掌握使用 SPSS Modeler 进行回归分析的基本方法
  - 4.1.4 学会使用 Logistic 节点进行回归分析
- 4.2 研究方法：根据电信客户流失的历史数据建立客户属性、服务属性、客户消费数据与客户流失可能性关联的数学模型，找出客户属性、服务属性、客户消费数据与客户流失最终状态的关系。
- 4.3 定义数据源
- 4.4 数据理解
- 4.5 数据准备
- 4.6 建立模型
  - 4.6.1 模型定义
  - 4.6.2 模型应用
- 4.7 建模步骤
  - 4.7.1 确定变量
  - 4.7.2 建立预测模型
  - 4.7.3 进行相关分析
  - 4.7.4 计算预测误差
  - 4.7.5 确定预测值
  - 4.7.6 根据模型进行预测、评估和分析
- 4.8 数据分析与模型评估
- 4.9 研究结论

## 模块四：实战演练篇—SPSS Modeler 算法及建模实战演练与操作

### 1. 实战演练项目

- 1.1 实战项目一：分类分析—4G 套餐类型与客户细分
- 1.2 实战项目二：聚类分析—4G 客户行为分析
- 1.3 实战项目三：关联分析—4G 产品关联分析
- 1.4 实战项目四：因子分析—提取影响电信客户流失的主成分分析
- 1.5 实战项目五：判别分析—电信客户群判别分析
- 1.6 实战项目六：贝叶斯网络分析—4G 手机主动营销分析
- 1.7 实战项目七：神经网络—电信客户流失预测分析
- 1.8 实战项目八：社交网络分析—客户流失预警分析

### 2. 要求与方法

- 2.1 分组交流与讨论，每组自选一个项目进行演练；
- 2.2 可以结合工作实际，自定实战项目；
- 2.3 研究内容及练习数据材料各小组自行准备；
- 2.4 按照数据录入、定义数据源、数据理解、数据准备、探索分析、数据建模、模型评估、研究结论八大步骤，运用 SPSS Modeler 14.1 版本软件进行实战演练。

## 模块五：高级进阶篇—运用 SPSS Modeler 关联分析法提升 4G 网络满意度

### 1. 运用 SPSS Modeler 数据挖掘方法提升 4G 网络满意度的总体思路

- 2. 目标与要求**
  - 2.1**理解关联分析法的基本理论
  - 2.2**掌握关联分析法的建模步骤
  - 2.3**掌握使用 SPSS Modeler 进行关联分析的基本方法
  - 2.4**学会使用 Apriori 算法进行关联分析
- 3. 研究方法：运用关联分析方法分析 4G 网络故障的产生与 KPI 指标之间的关联性**
- 4. 数据采集方法与平台搭建**
  - 4.1**4G 网络数据输入-数据采集方法
    - 4.1.1** 硬采集方式
    - 4.1.2** 软采集方式
    - 4.1.3** 硬采集与软采集方式对比
  - 4.2**TD-LTE 网络数据采集系统架构
    - 4.2.1** 数据接入采集层
    - 4.2.2** 协议处理层
    - 4.2.3** xDR 数据处理层
    - 4.2.4** 呈现层
  - 4.3**TD-LTE 数据采集系统存储应用设备
    - 4.3.1** 存储设备
    - 4.3.2** 接口服务器
    - 4.3.3** 应用服务器
    - 4.3.4** 网管服务器
    - 4.3.5** 数据库服务器
  - 4.4**TD-LTE 网络主要采集接口与协议
    - 4.4.1** S1-C 接口：S1AP、NAS 协议
    - 4.4.2** S1-U 接口：GTPv1-U、SIP、HTTP 等
    - 4.4.3** S5/S8 接口：GTPv2-C、GTPv1-U、PMIP
    - 4.4.4** S6a 接口：Diameter
    - 4.4.5** S3 接口：GTPv2-C
    - 4.4.6** S4 接口：GTPv2-C、GTPv1-U
    - 4.4.7** S10 接口：GTPv2-C
    - 4.4.8** S11 接口：GTPv2-C
    - 4.4.9** SGi 接口：DHCP、SIP、HTTP、Diameter 等
    - 4.4.10** Gx 接口：Diameter
    - 4.4.11** Rx 接口：IP、Diameter
    - 4.4.12** X2 接口：X2AP、GTPv1-U
- 5. 数据智能化预处理及关联分析**
  - 5.1**容量管理算法及流程
  - 5.2**质量分析及流程
    - 5.2.1** 栅格分析
    - 5.2.2** 与 DT&CQT 关联分析
  - 5.3**与容量的关联分析
  - 5.4**异常事件分析流程
  - 5.5**用户感知评估流程
- 6. 数据理解**

- 7. 数据准备**
- 8. 建立模型**
  - 8.1 模型定义**
  - 8.2 模型应用**
- 9. 数据挖掘输出-网络方面**
  - 9.1 网络质量/容量评估**
  - 9.2 异常事件**
- 10. 数据挖掘输出-用户与网络方面**
  - 10.1 用户行为刻画**
  - 10.2 网络感知评估**
  - 10.3 建立网络健康档案库**
- 11. 应用数据挖掘输出结果进行网络质量提升**
  - 11.1 网络质量评估**
  - 11.2 网络质量、容量预警**
  - 11.3 网络问题精准定位**
    - 11.3.1 PDN 连接 (时延/成功率/失败原因)**
    - 11.3.2 附着 (时延/成功率/失败原因)**
    - 11.3.3 服务请求 (时延/成功率/失败原因)**
    - 11.3.4 寻呼用户 (时延/成功率/失败原因)**
    - 11.3.5 缺省/专用承载建立 (时延/成功率/失败原因)**
    - 11.3.6 SIP 连接建立 (时延/成功率/失败原因)**
    - 11.3.7 专用承载建立**
    - 11.3.8 VoIP 通过建立**
- 12. 应用数据挖掘输出结果进行网络满意度提升**
  - 12.1 决策依据**
  - 12.2 用户感知评估**
  - 12.3 投诉预处理**
  - 12.4 客户关怀**
  - 12.5 用户行为分析**
- 13. 网络满意度评估-网络评估**
  - 13.1 质量评估**
  - 13.2 容量评估**
  - 13.3 异常事件原因分析**
  - 13.4 区域性/系统性问题精确定位**
- 14. 网络满意度评估-感知评估**
  - 14.1 感知评价体系建立**
  - 14.2 用户行为刻画**
  - 14.3 健康档案库建立**
- 15. 网络满意度评估-网络预警**
  - 15.1 话务、业务、流量模型建立**
  - 15.2 质量预警：质量突变、呼吸效应**
  - 15.3 容量预警：板件、无线资源**
  - 15.4 网络规划建议**
- 16. 网络满意度评估-服务支撑**

- 16.1 决策依据
- 16.2 投诉处理分析
- 16.3 客户关怀
- 16.4 行为分析、市场支撑

## 17.研究结论

## 18.形成分析报告

### 模块六：撰写分析报告篇一如何撰写漂亮的数据分析报告

#### 1. 数据分析思路与方法及案例分析

- 1.1 PEST 分析法与案例分析
- 1.2 5W2H 分析法与案例分析
- 1.3 逻辑树分析法与案例分析
- 1.4 4P 营销理论与案例分析
- 1.5 用户使用行为理论与案例分析

#### 2. 数据分析的基本方法

- 2.1 定量分析
- 2.2 定性分析
- 2.3 战略分析
- 2.4 竞争和客户分析
- 2.5 环境分析
- 2.6 评估分析
- 2.7 财务分析
- 2.8 宏观环境分析
- 2.9 竞争定位分析
- 2.10 竞标比超模式
- 2.11 情景分析
- 2.12 内容分析
- 2.13 竞争预警

#### 3. 数据分析报告的撰写

##### 3.1 撰写数据分析报告总体概述

3.1.1 从公开资料及互联网中搜集信息的方法

3.1.2 撰写数据分析报告的工作流程

##### 3.2 数据分析报告的撰写

3.2.1 分析报告的结构和要求

3.2.2 分析报告任务的组织实施：实施流程

3.2.3 分析报告的编写：内容结构

##### 3.3 态势分析报告的撰写

##### 3.4 评价分析报告的撰写

##### 3.5 预测分析报告的撰写

**3.6** 撰写报告对人员基本素质的要求

**4.** 演示部分优秀的分析报告

## **六、 课程总结：**

一、重点知识回顾与总结

二、互动与讨论：问与答

就学员提出的问题进行分析、讨论、模拟演练和点评。

## **七、 讲师介绍：**