

# CDA 数据分析员培训

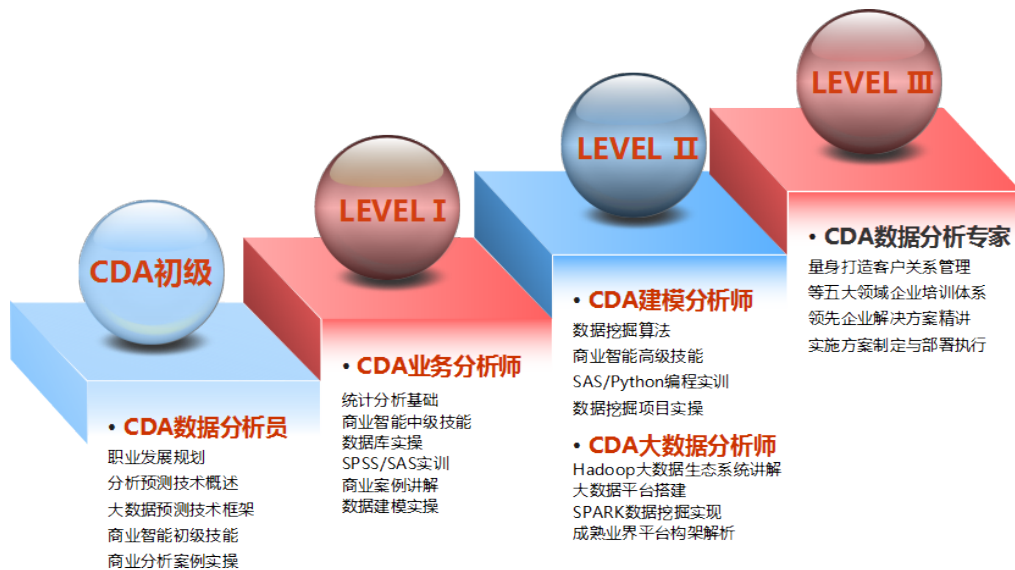
## 【授课对象】

本课程要求有一定的数学基础，掌握概率论和统计理论基础，有一定的数据分析经验。

适合于如下人员：数据分析师、大数据系统研发人员、大数据系统架构师、业务支撑部等对业务数据分析与挖掘有专业要求的相关人员。

## 【课程目标】

CDA（Certified Data Analyst）全称“注册数据分析师”，由“CERTIFIED DATA ANALYST INSTITUTE（CDA 数据分析师协会）”发起成立的职业简称，旨在培养正规化、科学化、专业化的数据分析人才队伍，进一步提升数据分析师的职业素养与能力水平，促进数据分析行业的高质量持续快速发展。CDA 数据分析师面向商业职场数据分析，分为三个等级，在国内由人大经济论坛主办资格考试，通过考试者可以获得 CDA 协会颁发的数据分析师资格证书，此证书代表数据分析师人才技能水平，为企业事业单位选拔和聘用专业人才的参考依据。



本课程基于 CDA 的认证知识要求，围绕商业问题，全面介绍了数据挖掘的标准流程，数据预处理，数据挖掘方法，数据挖掘模型，模型评估，模型参数优化，等等，使得学员掌握数据挖掘的方法、模型、工具。

通过本课程的学习，达到如下目的：

- 1、了解数据分析与数据挖掘的基本知识，理解大数据思维方式。
- 2、掌握数据挖掘的基本过程和步骤，掌握数据挖掘的思路和框架。
- 3、能够理解分析模型原理，掌握模型应用场景，能够利用模型解决复杂的商业问题。
- 4、掌握常用的数据模型，能够根据商业问题选择合适的分析模型。
- 5、熟悉 SPSS 基本操作，掌握分析操作，能够解读分析结果，并转化为业

务策略。

### **【授课时间】**

4~5 天时间

### **【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 Excel 2013 软件。
- 3、 便携机中事先安装好 SPSS v19 软件。

注：讲师可以提供试用版本软件及分析数据源。

### **【授课方式】**

基础知识精讲 + 案例讲解 + 操作演练 + 实际业务问题分析 + SPSS 实际操作

本课程突出数据挖掘的实际应用，结合行业的典型应用特点，从实际问题入手，引出相关知识，进行大数据的收集与处理；引导学员思考，构建分析模型，进行数据分析与挖掘，以及数据呈现与解读，全过程演练操作，以达到提

升学员的数据综合分析能力，支撑运营决策的目的。

## 【课程大纲】

### 第一部分：数据集基础知识（了解你的数据集）

- 1、 数据集概述
- 2、 数据集的类型
- 3、 数据集属性的类型
  - 标称
  - 序数
  - 度量
- 4、 数据质量三要素
  - 准确性
  - 完整性
  - 一致性
- 5、 数据预处理的内容
  - 数据清理（缺失值、离群值的处理方法）
  - 数据归约（维灾难、维归约、主成分分析）
  - 特征子集选择

- 特征创建/属性构造
- 数据离散化和二元化
- 属性/变量转换

## 6、 数据探索性分析

- 统计汇总
- 可视化

演练：描述性分析（频数、描述、探索、分类汇总）

## 第二部分：数据挖掘流程（基础，决定你的高度）

### 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：4G 终端营销项目挖掘过程分析

## 案例：客户匹配度模型—数据建模

### 3、常用数据统计指标

- 集中程度：均值、中位数、众数
- 离开程度：方差、标准差、极差
- 分布趋势：偏度、峰度
- 理解分布：正态分布、T分布、F分布

### 4、SPSS 基本操作（预处理）

- 数据导入
- 数据排序（排序个案）
- 重复数据处理（标识重复个案）
- 缺失值处理（替换缺失值）
- 生成新变量（计算变量、重新编码）
- 数据分组（分类汇总）
- 数据合并（合并文件）

## 演练：SPSS 基本操作

## 第三部分：数据挖掘实战篇

### 1、参数检验分析（样本均值检验）

商业问题：如何验证营销效果的有效性？

- 参数检验概述
  - ◇ 单样本 T 检验
  - ◇ 两独立样本 T 检验
  - ◇ 两配对样本 T 检验

➤ 参数检验原理以及步骤

➤ 参数检验适用场景

案例：电信运营商的 ARPU 值评估分析（单样本）

案例：信用卡消费金额评估分析（单样本）

案例：营销方式与产品销量的影响分析（两独立样本）

案例：吸烟与胆固醇升高的分析（两独立样本）

案例：营销方式有效性评估（两配对样本）

案例：减肥效果评估（两配对样本）

## 2、非参数检验分析（样本分布检验）

商业问题：这些属性数据的分布情况如何？如何从数据分布中看出问题？

- 非参数检验概述
  - ◇ 单样本检验

◇ 两独立样本检验

◇ 两相关样本检验

◇ 两配对样本检验

➤ 非参数检验原理

➤ 卡方检验、二项分布、游程检验的原理及适用场景

案例：死亡分布检验（单样本-卡方检验）

案例：产品合格率检验（单样本-二项分布）

案例：身高分布差异检验（单样本-KS 检验）

案例：设备正常工作检验（单样本-随机分布）

案例：制造工艺差异检验（两独立样本-MW 检验）

案例：训练新方法有效性检验（两配对样本-符号/秩检验）

案例：促销方式效果检验（多相关样本-Friedman 检验）

案例：客户满意度差异检验（多相关样本-Cochran Q 检验）

案例：评委评判黑幕检验（多相关样本-Kendall W 检验）

### 3、相关分析（相关程度计算）

商业问题：这两个属性是否会相互影响？影响程度大吗？

➤ 相关分析概述

➤ 计算相关系数的三个公式

案例：家庭生活开支的相关分析（Pearson 简单相关系数）

案例：营销费用与销售额的相关分析

案例：哪些因素与汽车销量有相关性

案例：腰围与体重的相关分析（偏相关分析）

#### 4、方差分析（影响因素分析）

商业问题：哪些才是影响销量的关键因素？

➤ 方差分析原理

➤ 方差分析的步骤

➤ 方差分析适用场景

➤ 如何解读方差分析结果

案例：终端陈列位置对终端销量的影响分析（单因素）

案例：广告形式、地区对销量的影响因素分析（多因素）

案例：2015 年大学生工资与父母职业的关系

案例：饲料对生猪体重的影响分析（协方差分析）

#### 5、回归分析（预测分析）

商业问题：如何预测未来的销售量（定量分析）？

- 回归分析概述及适用场景
- 回归分析的检验过程
- 如何选择最优回归模型
- 解读回归分析结果

案例：推广费用、办公费用与销售额的关系（线性回归）

案例：人均现金消费支出对人均食品消费支出的影响（曲线回归）

- 带分类变量的回归分析
- 比如，如何预测随着季节性变化的销量情况

案例：员工工龄、性别与终端销售的关系分析

案例：产品销量的季节性变化预测

## 6、逻辑回归分析（预测分析）

商业问题：如果评估用户购买某产品的概率？

- 逻辑回归分析原理
- 逻辑回归分析的适用场景

案例：客户购买预测分析（二元逻辑回归）

案例：品牌选择预测分析（多项逻辑回归）

## 7、时间序列分析（预测分析）

商业问题：随着时间变化，未来的销量变化趋势如何？

➤ 时序分析概述

- ◇ 移动平均 MA 模型
- ◇ 指数平滑 ES 模型
- ◇ 自回归滑动平均 ARIMA 模型
- ◇ 季节分解模型

➤ 时序分析适用场景

案例：汽车销量预测分析（指数平滑）

案例：上交所指数收益率预测分析（ARIMA）

案例：服装销售数据季节性趋势预测分析（季节分解）

## 第四部分：高级数据挖掘方法

### 1、聚类分析（Clustering）

商业问题：我们的客户有几类？各类特征是什么？

- 聚类方法原理介绍
- 聚类方法适用场景
- 系统聚类（层次聚类）算法原理
- 如何判定最佳聚类类别数量

案例：数据聚类分析--小康指数划分（Q型聚类）

案例：变量聚类分析--裁判标准一致性分析（R型聚类）

➤ K均值聚类（快速聚类）算法原理

案例：移动三大品牌细分市场合适吗？

案例：如何评选优秀员工（固定聚类中心）

演练：如何选择新产品试销地点？

## 2、决策树分类分析（Classification）

商业问题：这类客户有什么特征？有什么潜在销售机会？

➤ 决策树原理介绍

➤ 构建决策树的三个关键问题

◇ 如何选择最佳属性来构建节点

◇ 如何分裂变量

◇ 修剪决策树

➤ 选择最优属性

◇ 熵、基尼索引、分类错误

◇ 属性划分增益

➤ 如何分裂变量

- ◇ 多元划分与二元划分
- ◇ 连续变量离散化 (最优划分点)

➤ 修剪决策树

- ◇ 剪枝原则
- ◇ 预剪枝与后剪枝

➤ 如何评估分类性能

案例：银行低信用客户特征分析 (决策树分类)

### 3、基于规则的分类

- 基于规则分类原理介绍
- 评估规则的质量
- 构建分类规则：顺序覆盖法
- 规则增长策略

### 4、最近邻分类

### 5、朴素贝叶斯分类

- 贝叶斯分类原理
- 计算类别属性的条件概率
- 估计连续属性的条件概率

- 预测分类概率 (计算概率)

案例：评估银行用户拖欠贷款的概率

## 6、人工神经网络 (ANN)

- 神经网络基本原理
- 神经网络的结构
- ANN 关键问题
- MLP 与 RBF

案例：评估银行用户拖欠贷款的概率

## 7、判别分析

- 判别分析原理
- 距离判别法
- 典型判别法
- 贝叶斯判别法

案例：MBA 学生录取判别分析

案例：上市公司类别评估

## 8、关联分析 (Association)

商业问题：购买 A 产品的顾客还常常要购买其他什么产品？

- 关联规则原理介绍
- 关联规则的两个关键参数
  - ◇ 支持度
  - ◇ 置信度
- Apriori 算法介绍
- FP-Growth 算法介绍
- 关联规则适用场景

案例：商品套餐设计、商品交叉销售（关联分析）

## 9、客户价值评估 RFM 模型

商业问题：如何评估客户的价值？不同价值客户的营销策略有什么区别？

- RFM 模型介绍
- RFM 模型用户分类与业务策略
- RFM 与客户活跃度分析

案例：客户价值如何评估（什么才是 VIP 用户）

案例：如何选择促销用户（响应模型与促销）

案例：回头客用户特征分析（决策树分析）

## 10、主成分分析

- 主成分分析方法介绍
- 主成分分析基本思想
- 主成分分析步骤

案例：评估汽车购买者关注的哪些因素

## 第五部分：统计图表篇（看图说话）

- 1、柱状图/线图/饼图/高低图/箱图/散点图/直方图
- 2、图形的表达及适用场景

案例：各种图形绘制

**实战 1：客户流失预警与客户挽留之真实数据分析实践**

**实战 2：银行信用风险分析**

**结束：课程总结与问题答疑。**