

# 大数据挖掘及分析培训

## 一、课程目标：

本次课程以理论与实际相结合为基准，突出实际性演练，以达到如

## 下二、培训目标：

- (1) 了解数据挖掘的基本概念，基本流程、常用算法和适用的场景。
- (2) 能根据实际问题熟练地利用构建数据挖掘项目和流程，熟悉掌握数据建模过程、处理节点的操作。
- (3) 熟悉掌握具体的大数据统计分析技术以及案例，实战项目的具体应用。

## 三、培训对象：

- (1) 从事企业数据挖掘及相关工作的决策分析、工程技术人员；
- (2) 需要进行大数据统计分析的高校、科研院所的科技工作者；
- (3) 打算从事大数据统计分析的在校生、在职人员。

## 四、课程大纲：

### 一、大数据挖掘及分析技相关概念

#### 1、大数据挖掘及分析技基本概念

1) 基本概念

2) 通用系统架构

2、大数据背景下的挖掘及分析发展

1) 数据仓库发展

2) 数据挖掘发展

3) 报表技术发展

二、数据挖掘及分析理论及工具

1、数据挖掘及分析论体系

2、数据仓库技术

1) 数据仓库技术发展

2) 数据仓库建设过程

3、数据挖掘技术

1) 数据挖掘相关算法

2) 聚类算法

3) 分类算法

4) 关联算法

5) 预测算法

6) 数据挖掘新技术

4、可视化体系与工具

1) 可视化体系

2) 相关工具

3) 可视化开发过程

三、数据挖掘的主要方法及工具

1、数据挖掘主要方法

2、决策树分类

3、神经网络

4、回归方法

5、聚类分析

6、数据挖掘方法比较

7、分类器的评估与选择

8、流行数据分析平台及数据挖掘工具介绍

四、数据挖掘建模过程

1、数据挖掘流程概述

1) 问题识别

2) 数据理解

3) 数据准备

4) 建立模型

5) 模型评价

6) 部署应用

2、 离群点发现

1) 基于统计的离群点检测

2) 基于距离的离群点检测

3) 局部离群点算法

4) 不平衡数据级联算法

五、 实战项目——交叉销售

1、 背景介绍

2、 案例数据展示及分析

3、 数据展示

4、 业务目标及分析要求

5、 数据挖掘过程

6、 数据预处理

7、划分数据集及生成目标变量

8、生成衍生变量

9、生成挖掘表

10、建立打分模型

六、大数据的统计分析

1、相关分析

2、回归分析

3、决策树分析

1) 决策树的基本概念

2) 分类回归树

3) 决策树的剪枝

4) 假设检验

5) 聚类分析

6) 预测分析

7) 单因子方差分析

七、大数据的统计分析案例

1、小概率的力量：航空安全与彩民信心

- 2、夜空中的灾难
- 3、次灾难惊人的巧合
- 4、27000年才有一次的中奖机会
- 5、白点黑点，换个角度看数据
- 6、在整个背景下评价数据
- 7、精心选择的数字更丰富