**大数据分析与可视化技术应用实战培训班**

地点：上海

培训收益

通过此次课程培训，可使学习者获得如下收益：

1.理解数据分析及R语言；

2.理解并掌握数据结构；

3.掌握数据导入导出的方法；

4.掌握基本数据管理方法；

5.理解并掌握数据探索；

6.掌握如何进行描述性统计分析；

7.掌握R语言绘图基础；

8.掌握聚类分析；

9.掌握关联规则；

10.掌握KNN近邻分类；

11.通过案例掌握用户行为挖掘算法；

12.掌握C5.0算法原理；

13.掌握CART算法原理；

14.理解并掌握Rattle工具使用及如何搭建数据分析平台。

培训特色

1. 本课程尽量避开数学公式，按照“讲清思想方法原理—结合具体案例—R语言实现细节”思路，让即使是几乎没有什么基础的学员，掌握数据挖掘和可视化的基本思路和模式，打下未来深入的良好基础，能在工作和学习中结合具体问题立马上手操作解决；

2. 课程注重学练结合的方法，会采取课中练习的方法，充分调动大家思考的积极性，在做中掌握相关知识和技能；

3. 课程紧紧抓住基于R语言的数据挖掘和可视化的重点和难点，详细的分析和讲解，在理解难、容易出错的地方反复提醒，以便学员在课后容易的进行自己复习和相关拓展。

日程安排

|  |  |  |
| --- | --- | --- |
| 时间 | 培训模块 | 培训内容 |
| 第一天  上午 | 数据处理及复杂数据可视化（一） | R语言入门及数据挖掘基本概念  R语言简介  常量、向量、矩阵、数组、数据框、列表、数据处理、编写程序、基本统计计算、R语言绘图基础、帮助及包的用法  R语言数据处理秘笈  介绍填、剪、揉太极三式的数据处理原理、方法、案例、技巧，以及Dplyr包、Tidyr包、Data.Table包的具体用法和代码讲解  数据挖掘基础  讲解数据挖掘的基本概念、典型问题，通过场景举例说明数据挖掘的用途；同时，结合数据挖掘标准流程讲解开展数据挖掘工作的各环节及主要技术。介绍数据挖掘常用算法及适合场景，比较不同数据挖掘工具的优缺点。 |
| 第一天  下午 | 数据处理及复杂数据可视化（二） | R语言数据挖掘与可视化分析  数据挖掘关键技术  详细讲解抽样、分区、样本平衡、特征选择、训练模型、评估模型的技术，并结合R语言讲解其代码实现。  Rattle数据挖掘工具  介绍Rattle在业界的使用情况，安装及具体使用方法，包括数据源、数据探索、统计检验、数据变换、聚类分析、关联挖掘、建模、评估模型以及日志功能。通过实际的数据演示各环节的具体操作过程，分析挖掘结果。  可视化分析  介绍可视化基本概念，可视化设计流程以及R语言可视化的优势；其次，介绍24种常见图表（拆线图、柱状图、散点图、雷达图、词云图、玫瑰图、桑基图、河流图等等）的适用场景和使用方法；最后，重点对Lattice、Rcharts、Recharts、Ggplot2包进行了详细讲解和案例演示。 |
| 第二天  上午 | 数据挖掘模型实战（一） | 关联规则算法原理与案例分析  算法讲解  关联规则概念、基本术语、Apriori算法实现原理、流程及手动演算案例；Arules包的介绍等等  案例：利用超市购物篮Groceries数据进行关联规则分析  介绍Groceries数据集、事务数据集以及常见操作；介绍数据转换为事务数据集的方法以及使用Apriori算法建立关联规则模型，同时，对分析关联规则结果的常用技术进行了讲解和演示。  聚类算法原理与案例分析  Kmeans算法讲解  介绍Kmeans聚类基本思路及算法流程；Kmeans函数介绍；数据标准化技术介绍  案例：对Iris数据集进行K均值聚类分析  介绍Iris数据集，并试图通过使用基本数值指标建立聚类分析模型，验证其与真实分类的区别  案例：对Mtcars数据集进行K均值聚类分析  介绍Mtcars数据集，并使用Kmeans算法对其进行聚类分析，以了解汽车的类别细分，同时给出合理解释并提炼判断规则  层次聚类算法讲解  介绍样本距离、类间距离的概念和计算方法，介绍层次聚类算法流程以及通过手动演算详细讲解层次聚类实现的过程；Hclust函数讲解  案例：对洛杉机街区数据进行层次聚类  介绍洛杉机数据，并使用层次聚类算法对其进行细分；通过细分进一步对数据进行了详细了解，并给出各类的详细描述与说明  案例：对哺乳动物的睡眼数据进行层次聚类  介绍哺乳动物的睡眠数据，使用层次聚类的方法对其进行细分，并给出各类的特点，并给出详细描述与说明 |
| 第二天  下午 | 数据挖掘模型实战（二） | KNN算法原理与案例分析  算法原理  介绍Knn基本概念、原理以及实现流程，同时给出Knn算法的优缺点和适用场景；详细介绍Kknn包  案例：对Iris数据集进行Knn分类  介绍使用Knn算法对Iris数据进行分类的方法、流程及代码演示，并对分类的结果进行分析  案例：对乳腺癌数据进行Knn分类  介绍乳腺癌数据集，并使用Knn算法对乳腺癌数据进行分类，同时给出实现的方法、流程及代码，并对分类效果进行分析  案例：对新闻文本数据进行Knn分类  介绍新闻文本数据，并使用Knn分类算法对新闻文本数据进行分类，同时给出实现的方法、流程及代码，并对分类效果进行分析  决策树原理与案例分析  介绍决策树基本概念、分类与回归问题的区别与联系、说明决策树建模的通用过程。  CART算法原理  介绍分类回归树的基本思路、算法实现流程，并通过案例说明其实现细节；介绍选择分割点的方法，拆分规则，代价复杂度以及Rpart包和Prune函数的介绍  案例：对Iris数据集运用CART算法进行分类  介绍使用CART算法对Iris数据进行分类的方法、原理和代码实现流程，并对分类效果进行比较分析  案例：对汽车数据运用CART对汽车重量进行预测  介绍使用CART算法对Mtcars数据中汽车的重量进行预测的方法、原理和代码实现流程，并对预测效果进行分析  C5.0算法原理  介绍ID3算法的算法实现原理和流程，同时讲解信息熵、后验熵、信息增益等基本概念；通过案例讲解ID3算法的具体过程以及ID3自身的问题；介绍C4.5算法的实现原理和流程，讲解增益比例的来源和优点，同时介绍C4.5对连续属性及有缺失样本的算法办法，此外，还将会对剪枝、规则的产生、交叉验证等问题进行讲解；C50包的介绍  案例：对Iris数据集运用C50算法分类  介绍使用C50算法对Iris数据集进行分类的方法、原理和代码实现流程，同时对分类效果进行分析 |
| 第三天  上午 | 行业应用案例分享（一） | 分析案例剖析  购物篮分析案例  Onlineretail数据集，包含用户购买商品的交易信息，本案例旨在从中挖掘出用户购买商品的关联规则规律，并分析各模式对应的含义及应用可行性。  航空公司客户价值分析案例  面对激烈的市场竞争，各航空公司都推出了更优惠的营销方式来吸引更多的客户，国内某航空公司面临着旅客流失、竞争力下降和航空资源未充分利用等经营危机。通过建立合理的客户价值评估模型，对客户进行分群，分析比较不同客户群的客户价值，并制定相应的营销策略  Lending Club信贷违约分析案例  Lending Club 信用贷款违约数据是美国网络贷款平台 Lendingclub 在2007-2015年间的信用贷款情况数据，主要包括贷款状态和还款信息。本例基于该数据对多维度信息进行详细的数据分析，主要结合词云、地图等分析方法，另外，针对贷款违约建立了预测模型，本节会介绍详细的建模过程及预测效果分析 |
| 第三天  下午 | 行业应用案例分享（二） | 用户行为挖掘算法及案例  推荐系统  介绍推荐系统的基本概念、以及常见推荐算法（包括基于人口统计学的推荐、基于内容的推荐、协同过滤推荐、基于关联规则的推荐、基于效用的推荐、基于知识推荐、基于标签的推荐、基于图的推荐等等），同时对不同的推荐算法进行比较；另外，针对构建推荐系统，给出了详细的说明；介绍了推荐系统评测的关键指标（包括用户满意度、准确度、覆盖率、多样性、惊喜度等等）；接着，对R语言中的Recommenderlab包的用法进行详细介绍  案例：基于Onlineretail数据建立推荐系统  介绍Onlineretail数据集，以及开展推荐系统搭建过程的方法及具体实现，案例中使用了RANDOM、UBCF、IBCF三种方法建立了预测模型，并给出了比较分析  序列模式挖掘  介绍序列模式挖掘的基本概念、术语、实现的思路及具体过程。主要介绍SPADE算法的原理以及其实现过程。针对具体的实现，详细介绍了R语言中的Arulessequences包  对Onlineretail数据集进行序列模式挖掘  基于Onlineretail数据集，使用Spade算法建立序列模式挖掘算法进行序列模式挖掘，并对挖掘出的结果进行解释分析 |
| 第四天 | 学习考核与业内经验交流 | |

授课专家

游老师 计算机硕士，大数据分析、挖掘、可视化专家，高级培训讲师，曾服务于华为技术有限公司等多家企业，专注于机器学习、数据挖掘、大数据、知识图谱等领域的研究、设计与实现，在互联网、电信、电力、军工等行业具有丰富的工程实践经验，对空间分析、欺诈检测、广告反作弊、推荐系统、客户画像、客户营销建模、知识抽取、智能问答、可视化分析、预测分析、系统架构、大数据端到端解决方案等方面具有深刻理解，多次作为R语言会议重要嘉宾出席会议并发表主题演讲，著有《R语言预测实战》等多本书籍。

谢老师 某集团上市公司数据分析部负责人，主要利用R语言进行大数据的挖掘和可视化工作。从事数据挖掘建模工作已有10年，曾经从事过咨询、电商、金融、电购、电力、游戏等行业，了解不同领域的数据特点。有丰富的利用R语言进行数据挖掘实战经验，部分研究成果曾获得国家专利。

培训费用

面授班培训费5800 元/人（含场地费、考试证书费、教材费、学习期间的午餐费） 食宿统一安排，费用自理。请学员带身份证复印件一张。

本课程由中国信息化培训中心颁发《大数据分析与可视化技术应用高级工程师》证书，证书查询网址：www.zpedu.org; 证书可作为专业技术人员职业能力考核的证明，以及专业技术人员岗位聘用、任职、定级和晋升职务的重要依据。

咨询电话：400-061-6586

1. **报名回执【大数据分析与可视化技术应用实战】**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 单位名称  （开发票名称） |  | | | | | |
| 快递地址 |  | | | | 邮编 |  |
| 联 系 人 |  | | 职位 |  | 电话 |  |
| Email |  | | 传真 |  | | |
| 学员姓名 | 身份证号  （做证书使用） | 邮箱 | 联系电话 | | 培训地点 | 是否住宿 |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
| 汇 款  方 式 | 户 名：北京中培伟业管理咨询有限公司  开户行：北京农村商业银行卢沟桥支行  帐 号：0203 0101 0300 0033 172 | | | | | 学员签字  或盖章 |
| 户 名：中培创成（北京）科技有限公司  开户行：建设银行北京生命园支行  帐 号：1100 1048 6000 5250 6592 | | | | |
| 户 名：北京高培伟业管理咨询有限公司  开户行：中国工商银行股份有限公司北京菜市口支行  帐 号：0200 0018 0920 0152 881 | | | | |