**大数据分析挖掘与可视化最佳实践**

**培训班**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **培训地点** | **北京** | **成都** | **武汉** | **青岛** | **成都** | **北京** |
| **培训时间** | **2月****28-3.2** | **4月****27-29** | **6月****17-19** | **8月****待定** | **10月****待定** | **12月****待定** |
| **培训方式** | **面授+直播** |
| **增值服务** | **1.凡参加网络直播课程的学员，后期同一课程的网络直播课程可以再免费学习一次。****2.凡参加网络直播的学员可以享受免费再参加一次本课程的线下面授教学。****3.视频回放免费学习一年。** |

1. **培训重点**
2. 数据分析实战
3. 数据挖掘理论及核心技术
4. 大数据算法原理及案例实现
5. Python应用实战
6. **培训特色**

1.理论与实践相结合、案例分析与行业应用穿插进行；

2.专家精彩内容解析、学员专题讨论、分组研究；

3.通过全面知识理解、专题技能和实践结合的授课方式。

1. **日程安排**

|  |  |  |
| --- | --- | --- |
| **时间** | **知识模块** | **授课内容** |
| **第一天****上午** | **数据分析实战** | 第一讲 零基础学Python讲解Python背景、国内发展状况、基础语法、数据结构及绘图操作等内容。特别针对向量计算这块，着重介绍Python在这方面的优势及用法。 |
| 第二讲 数据分析方法论讲解统计分析基础，包括统计学基本概念，假设检验，置信区间等基础，并结合数据案例说明其使用场景和运用方法。介绍数据分析流程和常见分析思路，并结合案例进行讲解。 |
| 第三讲 数据处理技法从数据接入、数据统计、数据转换等几个方面进行讲解。数据接入包含接入MySQL、Oracle、Hadoop等常见数据库操作；数据统计包含Pandas包的具体用法和讲解；数据转换包含对数据集的关联、合并、重塑等操作。此外，针对海量数据的情况下，介绍在Spark平台上的数据处理技术，并结合真实环境进行操作讲解。 |
| **第一天****下午** | **数据挖掘理论及核心技术** | 第四讲 认识数据挖掘讲解数据挖掘基本概念，细致讲解业务理解、数据理解、数据准备、建立模型、模型评估、模型部署各环节的工作内容及相关技术；结合业界经典场景，讲解数据挖掘的实施流程和方法体系。 |
| 第五讲 数据挖掘核心技术细致讲解抽样、分区、样本平衡、特征选择、训练模型、评估模型等数据挖掘核心技术原理，并结合案例讲解其具体实现和用法。尤其针对样本平衡，重点讲解人工合成、代价敏感等算法；针对特征选择，重点讲解特征选择的核心思路，并结合Python进行案例演示。 |
| **第二天****上午** | **大数据算法原理及案例实现（1）** | 第六讲 特征降维算法及Python实现降维是大数据分析非常重要的算法，它可以在降低极少信息量的情况下，极大地缩小数据规模。主要讲解主成分、LDA以及t-SNE原理，并结合案例进行Python实现。特别地，针对海量数据情况下的应用场景，讲解实现思路和Python案例。 |
| 第七讲 决策树算法及Python实现决策树是非常经典的算法，一般常见于小数据的挖掘。由于决策树具有极强的可解释性，针对海量数据仍然是非常重要的实用价值。主要讲解ID3、C4.5、C5.0以及CART决策树算法的实现原理，并结合案例进行Python实现。 |
| 第八讲 好莱坞百万级影评数据分析与电影推荐实现实战部分：基于好莱坞百万级的影评数据，对数据进行建模、清洗、透视表操作。然后根据用户画像分析不同的用户喜好通过机器学习算法对不同性别、年龄阶段的用户进行定制化的电影推荐，最后把推荐的电影进行可视化的展示操作 |
| **第二天****下午** | **大数据算法原理及案例实现（2）** | 第九讲 因果推理算法及Python实现大数据分析技术可以帮助我们去发现、解决一些业务问题，然而如何去判断我们的改进是否生效，是否在业务指标上呈现过一定的因果逻辑，则是一个重要问题和分析方向。本节主要介绍因果推理算法，包括贝叶斯推理、状态空间模型以及CausalImpact工具等内容，并结合案例进行Python实现。 |
| 第十讲 深度学习算法及Python实现对于大数据的建模任务，我们可以基于深度学习来实现，不仅能够针对海量数据进行建模，其效果也非常不错。本节主要讲解深度学习的发展历程，DBN、DNN等经典深度学习算法，深度学习优化算法以及一些技巧。同时，介绍Keras、OpenCV库的使用方法，并结合案例进行Python实现。 |
| 第十一讲 采用OpenCV实现计算机视觉技术实战部分：基于OpenCV面部模型，完成对图片和视频的人脸识别，实战中会讲解OpenCV的重要类和函数。主要内容包括OpenCV库的安装和部署、图像增强、像素操作、图形分析等各种技术，并且详细介绍了如何处理来自文件或摄像机的视频，以及如何检测和跟踪移动对象。 |
| **第三天****上午** | **Python应用实战（一）** | 第十二讲 Python自然语言处理原理及案例目前文档数据已经成为很多企业重要的资产，通过对文档数据进行解析、建模、分析、挖掘、可视化，我们能够发现不一样的洞察。本节主要讲解自然语言处理基本概念和技法，包含分词、关键字提取、文摘提取、文本分类、主题模型、word2vec等内容。介绍在深度学习的加持下，与传统做法的区别，并使用Python进行案例讲解。 |
| 第十三讲 数据分析图表及Python案例数据可视化是大数据分析的重要手段，通过合理地使用图表，不仅可以简洁地表达数据的含义，高效地发现问题，还可以为报告的编写以及数据分析web应用增色不少。本节主要讲解常用的数据分析图表及其使用场景，介绍数据可视化的方法论，避免生搬硬套的使用图表，针对不同的业务场景和需求，合理选择可视化方法。介绍的工具不限于matplotlib、pycha、pyecharts、ggplot、Bokeh、HoloViews、mpld3、plotly、pygal等常用可视化库。 |
| **第三天****下午** | **Python应用实战（二）** | 第十四讲 使用Notebook编写数据分析报告数据分析报告在大数据分析过程中具有重要价值，它体现了大数据分析的目的、过程和结果，以及对发现问题的解读、改进方案等等，本节主要讲解使用Notebook编写数据分析报告的具体方法，以及编写数据分析报告的方法论，并结合案例讲解其用法。 |
| 第十五讲 Seaborn可视化开发实战Seaborn是一款不错的可视化框架，它和 Pandas一样是建立在 Matplotlib 之上的。可以基于Seaborn快速开发一个轻量级的数据分析web应用。在网页中嵌入图表、数据以及分析的算法，非常适合打造企业内部的敏捷数据分析工具集。本节主要介绍Pie、Scatter、Radar等等各种可视化解决方案，同时讲解一个用Seaborn实现数据分析功能（兼图表）的实际案例，搭建服务器，在企业内部实现轻量级数据分析应用。 |

1. **授课专家**

**刘老师** 10多年的IT领域相关技术研究和项目开发工作，在长期软件领域工作过程中，对软件企业运作模式有深入研究，熟悉软件质量保障标准ISO9003和软件过程改进模型CMM/CMMI，在具体项目实施过程中总结经验，有深刻认识。通晓多种软件设计和开发工具。对软件开发整个流程非常熟悉，能根据项目特点定制具体软件过程，并进行项目管理和监控，有很强的软件项目组织管理能力。对C/C++、HTML 5、python、Hadoop、java、java EE、android、IOS、大数据、云计算有比较深入的理解和应用，具有较强的移动互联网应用需求分析和系统设计能力，熟悉Android框架、IOS框架等技术，了解各种设计模式，能在具体项目中灵活运用。

1. **培训费用**

面授班培训费7800 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。

本课程由中国信息化培训中心颁发《大数据分析高级工程师》证书，证书查询网址：www.zpedu.com;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。

1. **报名回执【大数据分析挖掘与可视化最佳实践】**

|  |  |
| --- | --- |
| 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（培训方式） | 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 | 备注： |
| 发票信息： | 单位名称：统一社会信用代码：开户银行：账号：地址：电话： |

1. **联系方式**

联系人：赵娟

手 机（微信）：13261046023

QQ：1164082541

邮 箱：1164082541@qq.com

二〇二二年一月一日