

全国高校大数据与人工智能师资研修班 邀请函

(2024 年第一期)

线下（上海班）：数据采集与机器学习实战

线下（广州班）：大数据技术应用实战（Hadoop+Spark）

线上班（十一大专题）：PyTorch 深度学习与大模型应用实战、数据采集与处理实战、大数据分析机器学习实战、大数据技术应用实战（Hadoop+Spark）、商务数据分析实战（Excel+Power BI）、商务数据分析实战（Python）、计算机视觉应用实战（Pytorch）、文本分析与挖掘实战（PyTorch）、网络舆情与情感分析（Python）、TensorFlow 与人工智能实战、Python 金融数据分析与实战

主办单位：泰迪杯数据挖掘挑战赛组织委员会

协办单位：人民邮电出版社有限公司

广州泰迪教育科技有限公司

北京泰迪云智信息技术研究院

承办单位：广东泰迪智能科技股份有限公司

新一代信息技术正深刻改变着人们的生产、生活、学习和思维方式，其朝气蓬勃的产业发展态势和大力度的国家政策鼓励支持，无不揭示了数据智能技术的重要性。人才的培养是新一轮科技较量的基础，高等院校承担着大数据与人工智能人才培养的重任，因此，尽快形成完善的人才培养方案、课程体系、系列教材、课程师资和实训平台等，是各大高校共同关注的重要事宜。

为推动各院校完善专业建设，解决专业教学中行业案例缺失和实战能力不足等相关痛点问题，泰迪科技基于十余年深耕数据智能产业实践经验，特推出全国高校大数据与人工

智能师资研修班，每年在全国范围内滚动开展，截止目前已在全国巡回举办 60 余场，参训教师近 7000 人次。2024 年第一期全国高校大数据与人工智能师资研修班将开设十一个线上专题、两个线下专题（上海&广州），现将有关安排通知如下。

一、课程特色

1、大模型应用：畅想数据智能新未来

紧贴产业前沿，特邀三位专家畅谈大模型应用，分享“ChatGPT 教学应用”，“金牌助手 ChatGPT，让应用开发更高效”，“大模型(ChatGLM2-6B) 产业应用漫步与畅想”等专题，一起畅想数据智能新未来。

2、企业案例实战+落地实操授课

课程全程强调动手实操，内容以代码落地为主，通过讲解企业级案例，真正让学员把所学内容和工作实际有效结合，从而更好地完成自身后续教育教学工作。

3、全方位答疑辅导学习

课程设有答疑交流讨论群，培训期间助教全程辅助教学。线上学习部分每天提供 10 小时的实时在线答疑辅导，线下学习部分则有主讲老师和助教面对面的帮助。

4、内容从浅及深更易入门

本课程配套有基础知识内容，即使零基础学员也能找到适合自己的学习内容和节奏，快速掌握课程知识和技能。

5、提供课程资源和回看功能

所有课程相关源代码、数据、PPT、案例素材全部提供下载，即学即用，教学更轻松！线上课程内容支持六个月内免费回看，以便复习和参考。

6、满足教学和科研需求

通过技能学习和案例实战，学员将在具体应用场景中全面掌握相关技能，提升实训教学工作的实际动手能力并为后续科研打下坚实基础。

二、课程安排

本次培训分为线下和线上两种班型，详细课程内容见文末附件课程大纲。

1 线下班（上海&广州）

混合模式，技能学习课程放置在线上云课堂以便学员系统学习，核心案例实战采用线下形式由讲师亲授同步操作演练。

<线下上海班> [数据采集与机器学习实战\(Python\)](#)

基本信息	技能学习(线上云课堂), 1.14-1.16	现场案例实战, 1.18-1.21
时间: 1月14日-21日	1. Python 编程基础	1. 核心技能串讲与巩固
线下报到时间: 1月17日	2. Python 数据分析与应用	2. 泰迪内推平台招聘信息采集与分析
地点: 上海(详细地点开课 前一周发送报到通知)	3. Python 网络爬虫实战	3. 某外卖平台优惠券使用预测
学时: 共计 80 学时	4. Python 机器学习实战	4. 实战演练: 天猫用户重复购买预测 (从理论到实践)
证书: 高级机器学习工程师 职业技术证书	5. Python 特征工程实战(拓展)	
费用: 3980 元/人	6. 农产品信息采集与分析(拓展)	
学习环境要求: Windows10 或以上操作系统(64位), 8G+内存	7. 特别内容: ① ChatGPT 教学应用 ② 金牌助手 ChatGPT, 让应用开发更 高效 ③ 大模型(ChatGLM2-6B)产业应用漫 步与畅想	

<线下广州班> [大数据技术应用实战\(Hadoop+Spark\)](#)

基本信息	技能学习(线上云课堂), 1.16-1.20	现场案例实战, 1.22-1.26
时间: 1月16日-26日	1. Linux 操作系统基础	1. 航空客户价值分析
线下报到时间: 1月21日	2. Java 编程基础	2. 广电大数据用户画像(Hadoop + Spark + Hive)
地点: 广州(详细地点开课 前一周发送报到通知)	3. Scala 编程基础	3. 大数据分布式消息 Zookeeper(拓 展)
学时: 共计 80 学时	4. Hadoop 大数据基础	4. 商品实时推荐系统(拓展)
证书: 高级大数据技术应用 职业技术证书	5. Hive 大数据仓库	
费用: 4980 元/人	6. Spark 大数据技术基础	
学习环境要求: Windows10 或以上操作系统(64位), 8G+内存, 不可使用 Mac 和 linux 系统	7. HBase 非关系型数据库(拓展) 8. Flume 数据采集(拓展) 9. Kafka 消息系统(拓展) 10. Flink 大数据实时处理(拓展) 11. 特别内容: ① ChatGPT 教学应用 ② 金牌助手 ChatGPT, 让应用开发更 高效 ③ 大模型(ChatGLM2-6B)产业应用漫 步与畅想	

2 线上班（十一大专题）

教学视频六个月有效期可反复观看学习，同步提供配套资源，线上学习无需脱产，可灵

活安排时间。

专题一 [PyTorch 深度学习与大模型应用实战](#)

基本信息	技能学习	案例实战
时间：1月16日-26日 学时：共计88学时 证书：高级人工智能应用工程师 职业技术证书 费用：2480元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 机器学习实战 4. PyTorch 框架基础实践 5. PyTorch 深度学习原理与实现 6. 大模型原理与实现 7. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 新冠疫情期间网民情绪识别 2. 基于Transformer的疫情问诊系统自动翻译实现 3. 基于大模型的金融问句语义相似度计算（拓展）

专题二 [数据采集与处理实战 \(Python\)](#)

基本信息	技能学习	案例实战
时间：1月16日-25日 学时：共计80学时 证书：高级Python技术应用工程师 职业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Pandas 数据分析基础 4. Python 网络爬虫实战 5. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 泰迪内推平台招聘信息采集与分析 2. 农产品信息采集与分析 3. 微博疫情评论数据采集 4. 网站图像素材采集实战（拓展）

专题三 [大数据分析 with 机器学习实战 \(Python\)](#)

基本信息	技能学习	案例实战
时间：1月16日-25日 学时：共计80学时 证书：高级机器学习工程师 职业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Pandas 数据分析基础 4. Python 机器学习实战 5. Python 特征工程实战（拓展） 6. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 运营商用户流失预测 2. 网络入侵用户自动识别 3. 天猫用户重复购买预测 4. 泰迪内推平台信息精准推荐应用

专题四 [大数据技术应用实战 \(Hadoop+Spark\)](#)

基本信息	技能学习	案例实战
------	------	------

时间：1月16日-25日 学时：共计80学时 证书：高级大数据技术应用 职业技术证书 费用：2980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存，不可使用Mac和 linux系统	1. Linux 操作系统基础 2. Java 编程基础 3. Scala 编程基础 4. Hadoop 大数据基础 5. Hive 大数据仓库 6. Spark 大数据技术基础 7. HBase 非关系型数据库（拓展） 8. Flume 数据采集（拓展） 9. Kafka 消息系统（拓展） 10. Flink 大数据实时处理（拓展） 11. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更 高效 ③ 大模型(ChatGLM2-6B)产业应用漫 步与畅想	1. 航空客户价值分析 2. 广电大数据用户画像（Hadoop + Spark + Hive） 3. 大数据分布式消息 Zookeeper（拓 展） 4. 商品实时推荐系统（拓展）
---	--	---

专题五 [商务数据分析实战 \(Excel+Power BI\)](#)

基本信息	技能学习	案例实战
时间：1月16日-25日 学时：共计80学时 证书：高级大数据分析师职 业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 不可使用Mac和linux系统	1. Excel 数据分析基础与实战 2. Power BI 数据分析与可视化 3. Python 编程基础（拓展） 4. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更 高效 ③ 大模型(ChatGLM2-6B)产业应用漫 步与畅想	1. 新零售智能销售数据分析 2. 餐饮企业综合分析 3. 财务分析在纳税评估中的应用 4. 学生校园卡消费行为分析

专题六 [商务数据分析实战 \(Python\)](#)

基本信息	技能学习	案例实战
时间：1月16日-26日 学时：共计88学时 证书：高级大数据分析师职 业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 数据分析实训 4. 商务数据分析概述 5. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更 高效 ③ 大模型(ChatGLM2-6B)产业应用漫 步与畅想	1. 牛刀小试-T公司运营数据分析与 业务决策 2. 回归分析-某市财政收入预测 3. 聚类分析-对某航空公司客户分群 4. 分类分析-运营用户流失判别 5. 关联规则-购物篮商品推荐 6. 协同过滤-新闻网站用户个性化推 荐

专题七 [计算机视觉应用实战 \(PyTorch\)](#)

基本信息	技能学习	案例实战
------	------	------

时间：1月16日-26日 学时：共计88学时 证书：高级人工智能应用工程师 职业资格证书 费用：2480元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 机器学习实战 4. PyTorch 框架基础实践 5. PyTorch 深度学习原理与实现 6. 计算机视觉实战 7. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 脑PET图像分析和疾病预测 2. 基于FaceNet的人脸智能识别 3. 基于YOLOX的农田害虫图像检测与识别
---	--	---

专题八 [文本分析与挖掘实战 \(PyTorch\)](#)

基本信息	技能学习	案例实战
时间：1月16日-26日 学时：共计88学时 证书：高级人工智能应用工程师 职业资格证书 费用：2480元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 机器学习实战 4. PyTorch 框架基础实践 5. PyTorch 深度学习原理与实现 6. 自然语言处理实战 7. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 基于textCNN的公众健康问句分类 2. 新冠疫情期间网民情绪识别 3. 推荐系统受众性别智能识别

专题九 [网络舆情与情感分析 \(Python\)](#)

基本信息	技能学习	案例实战
时间：1月16日-26日 学时：共计88学时 证书：高级大数据分析师 职业资格证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）	1. Python 编程基础 2. Python 数据分析与应用 3. 网络舆情与情感分析概述 4. Python 机器学习实战 5. Python 网络爬虫实战 6. Python 文本挖掘实战 7. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 微博疫情评论数据爬虫 2. 基于情感分析的疫情期间网民情绪识别 3. 网络问政平台数据爬虫 4. 智慧政务下的问政舆情分析

专题十 [TensorFlow与人工智能实战](#)

基本信息	技能学习	案例实战
------	------	------

时间：1月16日-25日 学时：共计80学时 证书：高级人工智能应用工程师 职业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 机器学习实战 4. TensorFlow2 框架基础实践 5. TensorFlow2 深度学习原理与实现 6. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 脑PET图像分析与疾病预测 2. 基于textCNN的公众健康问句分类 3. 基于FaceNet的人脸智能识别
---	--	--

专题十一 Python 金融数据分析与实战

基本信息	技能学习	案例实战
时间：1月16日-25日 学时：共计80学时 证书：高级大数据分析师 职业技术证书 费用：1980元/人 学习环境要求：Windows10 或以上操作系统（64位）， 8G+内存	1. Python 编程基础 2. Python 数据分析与应用 3. Python 数据分析基础 4. Python 数据分析实训 5. Python 机器学习实战 6. 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想	1. 上市公司综合评价 2. 信用卡高风险客户分析 3. 金融客户产品购买预测

三、证书颁发

学员经培训并考试合格后，可以获得由工业和信息化部教育与考试中心颁发的相应职业技术证书，证书可登录工业和信息化部教育与考试中心官网查询。



四、报名及联系方式

1、报名材料：报名申请表、身份证复印件、两寸近期正面免冠彩色半身证件照电子版(要求：

背景：白色，格式：JPG，大小：14-20K)。

2、本期研修班由广州泰迪教育科技有限公司收取费用并开具发票。

3、联系方式

联系人：曾老师

微信：antonia602501

电话：13246821827

邮箱：zengaizhi@tipdm.com



全国高校大数据与人工智能师资研修班报名申请表

单位名称									
部门/院系									
通讯地址									
发票抬头					发票内容				
纳税号					电子发票接收邮箱				
联系人			电 话			邮 箱			
以下表格中要求提供的信息为申报职业资格证书使用，请真实完整填写。									
姓 名	性 别	职 务	主要任 课专业	毕业院 校	最高学历	手机号码	电子邮箱	专题选择 (线上/线下)	是否住宿 (单间/标间)
费用支付 方式	1、电汇到指定账号。 2、扫码支付（报名后联系工作人员索要支付码）。 3、付款时请注明“大数据与人工智能研修班+单位或姓名”字样，方便查账备案。								
账户信息	账户名：广州泰迪教育科技有限公司 开户行：中国工商银行广州花城支行 账户号：3602028509201584794								
备 注	请将报名表发送至邮箱：zengaizhi@tipdm.com 联系人：曾老师 13246821827 微信：antonia602501								

附件一 线下班课表

<线下上海班> 数据采集与机器学习实战 (Python) 课程大纲

技能学习 (线上云课堂) 课程安排		
<p>时间: 报名成功后即可开始学习</p> <p>Python 编程基础</p> <ol style="list-style-type: none">1 认识 Python2 编写 Python 程序3 认识 Python 数据结构4 条件判断及分支语句5 使用 def 定义函数6 认识面向对象7 读取文件数据8 模块和第三方库 <p>Python 数据分析与应用</p> <ol style="list-style-type: none">1 Python 数据分析概述2 NumPy 数值计算基础3 Matplotlib 数据可视化基础4. Pandas 数据分析基础5. 使用 Pandas 进行数据预处理 <p>时间: 1.14 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 网络爬虫实战</p> <p>1 Python 爬虫环境与爬虫简介</p> <ol style="list-style-type: none">1.1 认识爬虫1.2 认识反爬虫1.3 配置 Python 爬虫环境 <p>2 网页前端基础</p> <ol style="list-style-type: none">2.1 认识网络信息传输过程2.2 认识 HTTP <p>3 简单静态网页爬取</p> <ol style="list-style-type: none">3.1 认识静态网页3.2 实现 HTTP 请求3.3 解析网页3.4 存储数据 <p>4 认识动态网页</p> <ol style="list-style-type: none">4.1 认识动态网页4.2 逆向分析爬取动态网页4.3 使用 Selenium 爬取动态网页 <p>拓展</p>	<p>时间: 1.15 (9:00-18:00)</p> <p>Python 机器学习实战</p> <ol style="list-style-type: none">1 机器学习绪论1.1 引言1.2 基本术语1.3 假设空间&归纳偏好 <p>2 模型评估与选择</p> <ol style="list-style-type: none">2.1 经验误差与过拟合2.2 评估方法2.3 性能度量2.4 性能度量 Python 实现 <p>3 回归分析</p> <ol style="list-style-type: none">3.1 线性回归基本形式3.2 线性回归模型算法实现 <p>4 决策树</p> <ol style="list-style-type: none">4.1 从女生相亲到决策树4.2 明天适合打球吗4.3 决策树拆分属性选择4.4 决策树算法家族4.5 泰坦尼克号生还者预测—数据预处理4.6 泰坦尼克号生还者预测—模型构建与预测 <p>时间: 1.16 (9:00-18:00)</p> <p>5 神经网络</p> <ol style="list-style-type: none">5.1 单个神经元介绍5.2 经典网络结构介绍5.3 神经网络工作流程演示5.4 如何修正网络参数-梯度下降法5.5 网络工作原理推导5.6 网络搭建准备5.7 样本从输入层到隐层传输的 Python 实现5.8 网络输出的 Python 实现5.9 单样本网络训练实现5.10 全样本网络训练实现	<p>拓展内容:</p> <p>特征工程</p> <ol style="list-style-type: none">1 特征工程介绍2 数据预处理3 特征构建4 特征选择5 降噪与特征转换6 特征学习 <p>农产品信息采集与分析</p> <ol style="list-style-type: none">1.1 背景与分析目标2.1 网页分析和爬虫思路2.2 省份链接获取2.3 获取省份名称2.4 确定翻页数目2.5 获取单页表格2.6 获取所有省份和页面的数据2.7 分布式爬取3.1 数据预处理3.2 数据指标提取3.3 省级以上部门审定数量分析3.4 水稻品种类型数量分析3.5 主要水稻类型的数量分析3.6 水稻母本分析3.7 主要审定公司分析 <p>4 总结</p> <p>特别内容:</p> <ol style="list-style-type: none">① ChatGPT 教学应用② 金牌助手 ChatGPT, 让应用开发更高效③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想

5 模拟登录 6 Scrapy 爬虫 7. 拓展：终端协议及爬取工具介绍	5.11 网络性能评价 5.12 sklearn 实现神经网络 拓展 6 支持向量机 7 聚类 8 集成学习	
--	--	--

现场案例实战课程安排

<p>时间：1.18 (8:30-17:00)</p> <p>核心技能串讲与巩固</p> <p>1 Python 爬虫环境与爬虫简介 2 网页前端基础 3 简单静态网页爬取</p> <p>泰迪内推平台招聘信息采集与分析</p> <p>1 背景与目标 2 数据采集 2.1 网页结构探索 2.2 定位一级页面数据地址 2.3 爬取及解析一级页面数据 2.4 提取一级页面字段 2.5 定位二级页面数据地址 2.6 爬取及解析二级页面数据 2.7 翻页爬取及数据保存 3 数据处理 3.1 读取已爬取完成的数据 3.2 数据预处理操作 4 分析与可视化 4.1 招聘岗位对学历要求分析 4.2 各行业的大数据招聘需求数量分析 4.3 不同类型公司的薪资待遇分析 4.4 小结</p>	<p>时间：1.19 (8:30-17:00)</p> <p>核心技能串讲与巩固</p> <p>1 机器学习绪论 2 机器学习模型构建 3 模型评估与选择 4 集成学习</p> <p>某外卖平台优惠券使用预测</p> <p>1.1 背景与挖掘目标 2.1 工程环境准备 2.2 缺失值处理和数据去重 2.3 数据分布探索 3.1 特征工程介绍 3.2 原始特征 3.3 用户相关特征 3.4 商家相关特征 3.5 优惠券相关特征 3.6 离散型特征处理 4.1 建模前的数据处理 4.2 模型构建技巧 4.3 模型构建 4.4 网格搜索 4.5 模型训练和评估 4.6 模型优化 4.7 模型应用 5 小结</p>	<p>时间：1.20 (8:30-17:00)</p> <p>天猫用户重复购买预测：从理论到实践</p> <p>1 背景与目标 2 数据探索 3 构建样本标签 4 特征构建 5 模型构建 6 模型性能评估 7 模型预测 8 结果提交</p> <p>时间：1.21 (9:00-17:00)</p> <p>1. 职业技术考试 2. 企业参观 跳转至线下班课程安排</p>
--	--	---

<线下广州班> Hadoop+Spark 大数据技术应用实战课程大纲

技能学习（线上）课程安排

<p>时间：报名成功后即可开始学习</p> <p>Linux 操作系统基础</p> <ol style="list-style-type: none"> Linux 概述 Linux 系统安装 Linux 基本命令 Linux Vi 编辑器 Linux Shell 编程 <p>Java 编程基础</p> <ol style="list-style-type: none"> 基础知识 面向对象 线程及异常处理 <p>Scala 编程基础</p> <ol style="list-style-type: none"> Scala 简介 Scala 安装配置 Scala 基础语法 函数 面向对象编程 文件读写 	<p>时间：1.18 (9:00-18:00)</p> <p>Hive 大数据仓库</p> <ol style="list-style-type: none"> Hive 简介 <ol style="list-style-type: none"> Hive 简介 Hive 安装配置 <ol style="list-style-type: none"> Hive 安装配置之 MySQL 数据库安装 Hive 安装配置之 Hive 安装 Hive 应用 <ol style="list-style-type: none"> Hive 表定义 <ol style="list-style-type: none"> 创建内部表与外部表 创建静态分区表和动态分区表 创建带数据的表和桶表 Hive 导入及导出 <ol style="list-style-type: none"> Select 查询基本用法 1 Select 查询基本用法 2 内置函数应用 关联查询 <p>拓展</p> <ol style="list-style-type: none"> Hive 开发 Hive 自定义函数 Hive 查询优化 	<p>时间：1.20 (9:00-18:00)</p> <ol style="list-style-type: none"> 配置 Spark IDEA 开发环境 <ol style="list-style-type: none"> 搭建 Spark 开发环境 Spark SQL 应用 <ol style="list-style-type: none"> Spark SQL 简介 Spark SQL 配置 从结构化数据文件创建 DataFrame 从外部数据库创建 DataFrame 从 RDD 创建 DataFrame 读取 Hive 表数据创建 DataFrame 读取学生成绩创建 DataFrame 常见 DataFrame API 操作 1 常见的 DataFrame 操作 2 常见的 DataFrame 操作 3 通过 DataFrame API 计算学生总分和平均分 保存 DataFrame 数据 保存学生成绩分析结果到 Hive DataSet 基础操作 统计商品销量
<p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Hadoop 大数据基础</p> <ol style="list-style-type: none"> Hadoop 简介 <ol style="list-style-type: none"> 大数据介绍 Hadoop 核心组件 Hadoop 生态环境 Hadoop 应用场景 Hadoop 集群安装与部署 <ol style="list-style-type: none"> 安装虚拟机 安装 Java Hadoop 集群部署模式 配置固定 IP SSH 无密码登录 配置 Hadoop 集群 Hadoop 集群配置参数 Hadoop 集群启动与监控 	<p>时间：1.19 (9:00-18:00)</p> <p>Spark 大数据技术应用</p> <ol style="list-style-type: none"> Spark 入门 <ol style="list-style-type: none"> Spark 入门 Spark 集群的安装配置 <ol style="list-style-type: none"> Spark 安装部署 Spark 安装部署实战 Spark 架构及原理 <ol style="list-style-type: none"> Spark 架构 Spark RDD 及 DAG 相关概念 Spark 编程基础 <ol style="list-style-type: none"> 创建 RDD RDD 常用算子之 transformation 算子 (1) RDD 常用算子之 transformation 算子 (2) RDD 常用算子之键值对 RDD 算子 RDD 常用算子之 action 算子 	<p>拓展内容：</p> <ol style="list-style-type: none"> HBase 非关系型数据库 Zookeeper 大数据分布式消息 Flume 数据采集 Kafka 消息系统 Flink 大数据实时处理 商品实时推荐系统 <p>特别内容：</p> <ol style="list-style-type: none"> ChatGPT 教学应用 金牌助手 ChatGPT，让应用开发更高效 大模型 (ChatGLM2-6B) 产业应用漫步与畅想
<p>时间：1.17 (9:00-18:00)</p> <ol style="list-style-type: none"> Hadoop 基础操作 <ol style="list-style-type: none"> Hadoop 安全模式 Hadoop 集群基本信息介绍 HDFS 常用 Shell 操作 MapReduce 常用 Shell 操作 		

<p>3.5 MapReduce 任务管理</p> <p>3.6 Yarn 资源管理与任务调度</p> <p>4. MapReduce 开发入门</p> <p>4.1 MapReduce 框架与设计构思</p> <p>4.2 MapReduce 开发环境搭建</p> <p>4.3 MapReduce WordCount 源码分析</p> <p>4.4 MapReduce API 分析</p> <p>4.5 MapReduce 统计每天访问次数</p> <p>4.6 MapReduce 按用户访问次数排序</p> <p>拓展</p> <p>5. MapReduce 编程进阶</p>	<p>4.6 文件读取与存储</p> <p>4.7 统计用户停留时间最长的基站</p>	
--	---	--

案例实战（线下）课程安排		
---------------------	--	--

<p>时间：1.22(8:30)-1.24(11:30)</p> <p>核心技能串讲与巩固</p> <p>1. Hadoop 大数据基础</p> <p>2. Hive 大数据仓库</p> <p>3. HBase 非关系型数据库</p> <p>4. Spark 大数据技术应用</p> <p>时间：1.24(14:00-17:00)</p> <p>航空客户价值分析</p> <p>1. 案例背景与挖掘目标</p> <p>2. 数据存储</p> <p>3. 数据探索与数据清洗</p> <p>4. 构建航空客户价值分析关键特征</p> <p>5. 航空客户分群</p> <p>6. 会员营销策略与建议</p>	<p>时间：1.25(8.30-17:00)</p> <p>广电大数据用户画像</p> <p>1. 项目背景与目标分析</p> <p>1.1 背景介绍</p> <p>1.2 目标分析</p> <p>1.3 系统架构介绍</p> <p>2. 数据存储</p> <p>2.1 数据介绍</p> <p>2.2 数据存储</p> <p>2.3 开发环境搭建</p> <p>3. 数据探索与清洗</p> <p>4. 用户画像标签计算</p> <p>4.1 用户画像简介</p> <p>4.2 消费内容</p> <p>4.3 消费等级</p> <p>4.4 产品名称</p> <p>4.5 入网程度&业务名称</p> <p>4.6 地区&语言偏好</p> <p>5. SVM 预测用户是否挽留</p> <p>5.1 SVM 介绍</p> <p>5.2 特征构建</p> <p>5.3 标签构建</p> <p>5.4 模型构建与评估</p>	<p>时间：1.26(9:00-17:00)</p> <p>1. 职业技术考试</p> <p>2. 企业参观</p> <p>跳转至线下班课程安排</p>
--	---	---

附件二 线上班课表

专题一 PyTorch 深度学习与大模型应用实战课程大纲

技能学习课程安排

时间：报名成功后即可开始学习

Python 编程基础

- 1 认识 Python
- 2 编写 Python 程序
- 3 认识 Python 数据结构
- 4 条件判断及分支语句
- 5 使用 def 定义函数
- 6 认识面向对象
- 7 读取文件数据
- 8 模块和第三方库

Python 数据分析与应用

- 1 Python 数据分析概述
- 2 NumPy 数值计算基础
- 3 Matplotlib 数据可视化基础
- 4 Pandas 统计分析基础
- 5 使用 Pandas 进行数据预处理

时间：1.16 (9:00-18:00)

专题讲座

计算机视觉技术及其应用

Python 机器学习实战

- 1 机器学习绪论
 - 1.1 引言
 - 1.2 基本术语
 - 1.3 假设空间&归纳偏好
- 2 模型评估与选择
 - 2.1 经验误差与过拟合
 - 2.2 评估方法
 - 2.3 性能度量
 - 2.4 性能度量 Python 实现
- 3 回归分析
 - 3.1 线性回归基本形式
 - 3.2 线性回归模型的 Python 实现
- 4 人工神经网络
 - 4.1 单个神经元介绍
 - 4.2 经典网络结构介绍
 - 4.3 神经网络工作流程演示
 - 4.4 如何修正网络参数-梯度下降法
 - 4.5 网络工作原理推导

时间：1.17 (9:00-18:00)

- 4.6 网络搭建准备
- 4.7 样本从输入层到隐层传输的 Python 实现
- 4.8 网络输出的 Python 实现
- 4.9 单样本网络训练实现
- 4.10 全样本网络训练实现
- 4.11 网络性能评价
- 4.12 实现神经网络算法
- 5.1 聚类分析概述
- 5.2 相似性度量
- 5.3 K-Means 聚类分析算法介绍
- 5.4 利用 K-Means 算法对鸮尾花进行聚类
- 5.5 聚类结果的性能度量
- 5.6 调用 sklearn 实现聚类分析
- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量机算法实现

时间：1.18 (9:00-18:00)

PyTorch 框架基础实践

- 1 PyTorch 简介
- 2 张量操作
 - 2.1 创建张量
 - 2.2 张量与数组相互转化
- 3 构建一个线性模型
 - 3.1 任务描述：构建线性模型
 - 3.2 读取数据
 - 3.3 构建初始模型及损失函数
 - 3.4 test-构建优化器
 - 3.4 构建优化器
 - 3.5 最小化方差（训练）
 - 3.6 执行多轮训练
 - 3.7 训练过程可视化
- 4 识别手写数字
 - 4.1 案例目标与流程
 - 4.2 加载数据
 - 4.3 加工数据
 - 4.4 模型结构介绍
 - 4.5 构建模型

时间：1.19 (9:00-18:00)

PyTorch 深度学习原理与实现

- 1 引言
- 2 卷积神经网络 CNN
 - 2.1 浅层神经网络的局限
 - 2.2 卷积操作
 - 2.3 卷积操作的优势
 - 2.4 池化及全连接
 - 2.5 高维输入及多 filter 卷积
 - 2.6 实现卷积操作
 - 2.7 将卷积结果可视化
 - 2.8 实现池化操作

时间：1.20 (9:00-18:00)

- 3 循环神经网络 RNN
 - 3.1 循环神经网络简介
 - 3.2 循环神经网络的常见结构
- 4 长短时记忆网络 LSTM
 - 4.1 LSTM 的三个门
 - 4.2 LSTM 三个门的计算示例
 - 4.3 实现 LSTM 操作
 - 4.4 LSTM 返回值解读
- 5 利用 LSTM 实现手写数字识别
 - 5.1 加载数据
 - 5.2 数据加工
 - 5.3 搭建循环神经网络
 - 5.4 模型配置
 - 5.5 模型训练
 - 5.6 模型性能验证

时间：1.21 (9:00-18:00)

大模型原理与实现

- 1 基础模型（大模型）简介
- 2 万模基座 Transformer

时间：1.22 (9:00-18:00)

- 3 生成式预训练模型 GPT
- 4 双向编码模型 BERT
- 5 Transformer 应用

特别内容

- ① ChatGPT 教学应用
- ② 金牌助手 ChatGPT，让应用开发更高效

	4.6 模型配置 4.7 模型训练 4.8 执行多轮训练 4.9 模型性能评估 4.10 保存模型 4.11 加载模型 4.12 模型应用	③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想
--	---	-------------------------------

案例实战课程安排

<p>时间：1.23 (9:00-18:00) 新冠疫情期间网民情绪识别</p> <ol style="list-style-type: none"> 1 背景与目标 2 数据准备 <ol style="list-style-type: none"> 2.1 数据基本介绍 2.2 词嵌入介绍 2.3 进行词向量训练 2.4 构建词向量矩阵 2.5 获取编码后的语料库 2.6 对各样本执行 padding 操作 2.7 拆分数据并将其转为模型所需格式 3 模型训练与性能验证 <ol style="list-style-type: none"> 3.1 Embedding 层介绍 3.2 定义 Embedding 层算子 3.3 定义 LSTM 层算子 3.4 定义全连接层算子 3.5 定义网络计算流程 3.6 模型构建及配置 3.7 模型训练 3.8 模型性能评估 4 小结 	<p>时间：1.24 (9:00-18:00) 基于 Transformer 的疫情问诊系统自动翻译实现</p> <ol style="list-style-type: none"> 1.1 机器翻译发展历程 1.2 任务目标及实现流程 <ol style="list-style-type: none"> 2.1 数据加载 2.2 数据拆分 2.3 预训练模型介绍 2.4 Tokenizer 介绍 2.5 Tokenizer 调用实现 3 预训练模型加载 <p>时间：1.25 (16:00-18:00)</p> <ol style="list-style-type: none"> 4.1 定义模型训练参数 4.2 定义数据收集器 <ol style="list-style-type: none"> 4.3.1 BLEU 值概念与计算 4.3.2 定义 BLEU 值 4.4 模型训练 5.1 模型推理 5.2 小结 	<p>拓展内容： 基于大模型的金融问句语义相似度计算</p> <ol style="list-style-type: none"> 1 案例背景及挖掘目标 2 语义文本相似度 <ol style="list-style-type: none"> 3.1 预处理：文件加载 3.2 预处理：错别字纠正 3.3 预处理：数据类型转化与长度统计 3.4 预处理：数据准备 3.5 加载预训练模型 3.6 定义损失函数 3.7 模型训练 3.8 模型预测 3.9 模型评估 4 小结 <p>时间：1.26 (16:00-18:00) 职业技术考试 跳转至线上上班课程安排</p>
---	---	---

专题二 数据采集与处理实战 (Python) 课程大纲

技能学习课程安排

<p>时间：报名成功后即可开始学习 Python 编程基础</p> <ol style="list-style-type: none"> 1 认识 Python 2 编写 Python 程序 3 认识 Python 数据结构 4 条件判断及分支语句 5 使用 def 定义函数 6 认识面向对象 7 读取文件数据 	<p>时间：1.18 (9:00-18:00) 3 使用 Pandas 进行数据预处理 <ol style="list-style-type: none"> 3.1 合并数据 3.2 清洗数据 3.3 标准化数据 3.4 转换数据 <p>时间：1.19 (9:00-18:00) Python 网络爬虫实战</p> <ol style="list-style-type: none"> 1 Python 爬虫环境与爬虫简介 </p>	<p>时间：1.21 (9:00-18:00) 5 模拟登录 <ol style="list-style-type: none"> 5.1 使用表单登录方法实现模拟登录 5.2 使用 Cookie 登录方法实现模拟登录 5.3 使用 Selenium 模拟登录 拓展 6 Scrapy 爬虫</p>
--	--	--

<p>8 模块和第三方库</p> <p>Python 数据分析与应用</p> <p>1 Python 数据分析概述</p> <p>2 NumPy 数值计算基础</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 数据分析基础</p> <p>1.1 掌握绘图基础语法与常用参数</p> <p>1.2 分析特征间关系</p> <p>1.3 分析特征内部数据分布与分散情况</p> <p>时间：1.17 (9:00-18:00)</p> <p>2 Pandas 统计分析基础</p> <p>2.1 Pandas 简介</p> <p>2.2 读写不同数据源的数据</p> <p>2.3 数据框与数据框元素</p> <p>2.4 转换与处理时间序列数据</p> <p>2.5 使用分组聚合进行组内计算</p> <p>2.6 创建透视表与交叉表</p>	<p>1.1 认识爬虫</p> <p>1.2 认识反爬虫</p> <p>1.3 配置 Python 爬虫环境</p> <p>2 网页前端基础</p> <p>2.1 认识网络信息传输过程</p> <p>2.2 认识 HTTP</p> <p>3 简单静态网页爬取</p> <p>3.1 认识静态网页</p> <p>3.2 实现 HTTP 请求</p> <p>3.3 解析网页</p> <p>3.4 存储数据</p> <p>时间：1.20 (9:00-18:00)</p> <p>4 认识动态网页</p> <p>4.1 认识动态网页</p> <p>4.2 逆向分析爬取动态网页</p> <p>4.3 使用 Selenium 爬取动态网页</p>	<p>6.1 认识 Scrapy</p> <p>6.2 通过 Scrapy 爬取基本页面信息</p> <p>6.3 通过 Scrapy 抓取跳转页面数据</p> <p>7 拓展：终端协议及爬取工具介绍</p> <p>拓展内容：</p> <p>网站图像素材采集实战</p> <p>1 思路介绍</p> <p>2 单个图片文件爬取</p> <p>3 获取一个页面所有图片网址</p> <p>4 保存所有图片</p> <p>5 翻页爬取更多数据</p> <p>6 PDF 文件规律及问题</p> <p>7 PDF 翻页刷新的网址规律</p> <p>8 获取当前页所有图片网址</p> <p>9 翻页刷新爬取所有图片</p> <p>10 图片拼接成 PDF 文件</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT, 让应用开发更高效</p> <p>③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想</p>
案例实战课程安排		
<p>时间：1.22 (9:00-18:00)</p> <p>泰迪内推平台招聘信息采集与分析</p> <p>1 背景与目标</p> <p>2 数据采集</p> <p>2.1 网页结构探索</p> <p>2.2 定位一级页面数据地址</p> <p>2.3 爬取及解析一级页面数据</p> <p>2.4 提取一级页面字段</p> <p>2.5 定位二级页面数据地址</p> <p>2.6 爬取及解析二级页面数据</p> <p>2.7 翻页爬取及数据保存</p> <p>3 数据处理</p> <p>3.1 读取已爬取完成的数据</p> <p>3.2 数据预处理操作</p> <p>4 分析与可视化</p> <p>4.1 招聘岗位对学历要求分析</p> <p>4.2 各行业的大数据招聘需求数量分析</p>	<p>时间：1.23 (9:00-18:00)</p> <p>农产品信息采集与分析</p> <p>1.1 背景与分析目标</p> <p>2.1 网页分析和爬虫思路</p> <p>2.2 省份链接获取</p> <p>2.3 获取省份名称</p> <p>2.4 确定翻页数目</p> <p>2.5 获取单页表格</p> <p>2.6 获取所有省份和页面的数据</p> <p>2.7 分布式爬取</p> <p>3.1 数据预处理</p> <p>3.2 数据指标提取</p> <p>3.3 省级以上部门审定数量分析</p> <p>3.4 水稻品种类型数量分析</p> <p>3.5 主要水稻类型被审定的数量分析</p> <p>3.6 水稻母本分析</p> <p>3.7 主要审定公司分析</p> <p>4 总结</p>	<p>时间：1.24 (9:00-18:00)</p> <p>微博疫情评论数据采集</p> <p>1 背景与目标</p> <p>2.1 评论结构分析</p> <p>2.2 数据接口分析</p> <p>3.1 微博页面接口分析</p> <p>3.2 评论数据接口分析</p> <p>3.3 评论回复数据爬取</p> <p>3.4 单页微博及评论数据爬取</p> <p>3.5 多线程爬虫</p> <p>4 小结</p> <p>时间：1.25 (16:00-18:00)</p> <p>职业技术考试</p> <p>跳转至线上课程安排</p>

4.3 不同类型公司的薪资待遇分析		
4.4 小结		

专题三 大数据分析 with 机器学习实战 (Python) 课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <ol style="list-style-type: none"> 1 认识 Python 2 编写 Python 程序 3 认识 Python 数据结构 4 条件判断及分支语句 5 使用 def 定义函数 6 认识面向对象 7 读取文件数据 8 模块和第三方库 <p>Python 数据分析与应用</p> <ol style="list-style-type: none"> 1 Python 数据分析概述 2 NumPy 数值计算基础 3 Matplotlib 数据可视化基础 <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 数据分析基础</p> <ol style="list-style-type: none"> 1 Pandas 统计分析基础 1.1 Pandas 简介 1.2 读写不同数据源的数据 1.3 数据框与数据框元素 1.4 转换与处理时间序列数据 1.5 使用分组聚合进行组内计算 1.6 创建透视表与交叉表 <p>时间：1.17 (9:00-18:00)</p> <ol style="list-style-type: none"> 2 使用 Pandas 进行数据预处理 2.1 合并数据 2.2 清洗数据 2.3 标准化数据 2.4 转换数据 	<p>时间：1.18 (9:00-18:00)</p> <p>Python 机器学习实战</p> <ol style="list-style-type: none"> 1 机器学习绪论 1.1 引言 1.2 基本术语 1.3 假设空间&归纳偏好 2 模型评估与选择 2.1 经验误差与过拟合 2.2 评估方法 2.3 性能度量 2.4 性能度量 Python 实现 3 回归分析 3.1 线性回归基本形式 3.2 线性回归模型实现 4 决策树 4.1 从女生相亲到决策树 4.2 明天适合打球吗 4.3 决策树拆分属性选择 4.4 决策树算法家族 4.5 泰坦尼克号生还者预测—数据预处理 4.6 泰坦尼克号生还者预测—模型构建与预测 <p>时间：1.19 (9:00-18:00)</p> <ol style="list-style-type: none"> 5 人工神经网络 5.1 单个神经元介绍 5.2 经典网络结构介绍 5.3 神经网络工作流程演示 5.4 修正网络参数-梯度下降法 5.5 网络工作原理推导 5.6 网络搭建准备 5.7 样本从输入层到隐层传输的 Python 实现 	<p>时间：1.20 (9:00-18:00)</p> <ol style="list-style-type: none"> 6 支持向量机 6.1 间隔与支持向量 6.2 对偶问题 6.3 核函数 6.4 软间隔与正则化 6.5 支持向量机算法的 Python 实现 7 聚类算法 7.1 聚类分析概述 7.2 相似性度量 7.3 K-Means 聚类分析算法介绍 7.4 利用 K-Means 算法对鸢尾花进行聚类 7.5 聚类结果的性能度量 7.6 调用 sklearn 实现聚类分析 <p>拓展</p> <ol style="list-style-type: none"> 8 集成学习 8.1 集成学习基本概念 8.2 并行集成算法-Bagging&RandomForest 8.3 串行集成算法-Boosting 算法流程 8.4 串行集成算法-Boosting 代码实现 8.5 Stacking 算法流程 8.6 Stacking 代码实现 <p>拓展内容：</p> <p>特征工程</p> <ol style="list-style-type: none"> 1 特征工程介绍 2 数据预处理 3 特征构建 4 特征选择 5 降噪与特征转换

	5.8 网络输出的 Python 实现 5.9 单样本网络训练实现 5.10 全样本网络训练实现 5.11 网络性能评价 5.12 调用 sklearn 实现神经网络	6 特征学习 特别内容： ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想
--	---	--

案例实战课程安排

<p>时间：1.21 (9:00-18:00)</p> <p>运营商用户流失预测</p> 1.1 背景与目标 1.2 案例思路分析 2.1 数据探索 2.2 数据去重及删除无关属性 2.3 用户分组及标签构建 2.4 提取用户基本信息和在网时长 2.5 处理合约是否有效 2.6 处理合约计划到期时间 2.7 其余变量处理 2.8 特征拼接及缺失值处理 2.9 数据保存 3.1 特征选择介绍 3.2 皮尔逊特征选择 3.3 处理样本类别不均衡问题 4.1 模型性能评估介绍 4.2 模型构建及性能评估	<p>时间：1.23 (9:00-18:00)</p> <p>天猫用户重复购买预测</p> 1.1 背景与挖掘目标 2.1 工程环境准备 2.2 缺失值处理和数据去重 2.3 数据分布探索 3.1 特征工程介绍 3.2 原始特征 3.3.1 用户相关特征：用户在平台的总交互次数 3.3.2 用户相关特征：用户最近一次购买距离第一次的时长 3.4.1 商家相关特征：商家被交互的数量 3.4.2 商家相关特征：商家的复购次数 3.5.1 用户和商家相关特征：用户在商家的交互次数 3.5.2 用户和商家相关特征：不同用户在不同商家购买率 3.6 离散型特征处理 4.1 建模前的数据处理 4.2 模型构建 4.3 模型训练和评估 4.4 模型应用 5 小结	<p>时间：1.24 (9:00-18:00)</p> <p>泰迪内推平台信息精准推荐应用</p> 1 背景与目标 2 目标分析 2.1 推荐思路分析 2.2 基于物品的协同过滤推荐介绍 3 工程实现 3.1 EB 工具登录及简介 3.2 创建空白工程 3.3 导入数据 3.4 筛选正文数据 3.5 字符替换及记录去重 3.6 划分训练集用户和测试集用户 3.7 构造训练集和测试集数据 3.8 构建模型 3.9 推荐及性能评估 4 小结
<p>时间：1.22 (9:00-18:00)</p> <p>网络入侵用户自动识别</p> 1 背景与目标 2 数据处理 2.1 读取数据 2.2 了解数据基本情况 2.3 哑变量处理 2.4 拼接特征并删除无关列 2.5 标签转化及预处理函数封装 3 模型训练与验证 3.1 认识集成学习算法 3.2 模型训练与性能验证 3.3 保存模型集训练数据结构 4 模型应用与评估 4.1 加载并处理待预测样本 4.2 模型应用及性能评估		<p>时间：1.25 (16:00-18:00)</p> <p>职业技术考试</p> 跳转至线上课程安排

专题四 Hadoop+Spark 大数据技术应用实战课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Linux 操作系统基础</p> <ol style="list-style-type: none"> Linux 概述 Linux 系统安装 Linux 基本命令 Linux Vi 编辑器 Linux Shell 编程 <p>Java 编程基础</p> <ol style="list-style-type: none"> 基础知识 面向对象 线程及异常处理 <p>Scala 编程基础</p> <ol style="list-style-type: none"> Scala 简介 Scala 安装配置 Scala 基础语法 函数 面向对象编程 文件读写 <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Hadoop 大数据基础</p> <ol style="list-style-type: none"> Hadoop 简介 <ol style="list-style-type: none"> 1.1 大数据介绍 1.2 Hadoop 核心组件 1.3 Hadoop 生态环境 1.4 Hadoop 应用场景 Hadoop 集群安装与部署 <ol style="list-style-type: none"> 2.1 安装虚拟机 2.2 安装 Java 2.3 Hadoop 集群部署模式 2.4 配置固定 IP 2.5 SSH 无密码登录 2.6 配置 Hadoop 集群 2.7 Hadoop 集群配置参数 2.8 Hadoop 集群启动与监控 <p>时间：1.17 (9:00-18:00)</p> <ol style="list-style-type: none"> Hadoop 基础操作 <ol style="list-style-type: none"> 3.1 Hadoop 安全模式 	<p>时间：1.18 (9:00-18:00)</p> <p>Hive 大数据仓库</p> <ol style="list-style-type: none"> Hive 简介 <ol style="list-style-type: none"> 1.1 Hive 简介 Hive 安装配置 <ol style="list-style-type: none"> 2.1.1 Hive 安装配置之 MySQL 数据库安装 2.1.2 Hive 安装配置之 Hive 安装 Hive 应用 <ol style="list-style-type: none"> 3.1.1 Hive 表定义 3.1.2 创建内部表与外部表 3.1.3 创建静态分区表和动态分区表 3.1.4 创建带数据的表和桶表 3.2 Hive 导入及导出 <ol style="list-style-type: none"> 3.3.1 Select 查询基本用法 1 3.3.2 Select 查询基本用法 2 3.3.3 内置函数应用 3.3.4 关联查询 <p>时间：1.19 (9:00-18:00)</p> <p>拓展</p> <ol style="list-style-type: none"> Hive 开发 Hive 自定义函数 Hive 查询优化 <p>时间：1.20 (9:00-18:00)</p> <p>Spark 大数据技术应用</p> <ol style="list-style-type: none"> Spark 入门 <ol style="list-style-type: none"> 1.1 Spark 入门 Spark 集群的安装配置 <ol style="list-style-type: none"> 2.1 Spark 安装部署 2.2 Spark 安装部署实战 Spark 架构及原理 <ol style="list-style-type: none"> 3.1 Spark 架构 3.2 Spark RDD 及 DAG 相关概念 Spark 编程基础 <ol style="list-style-type: none"> 4.1 创建 RDD 4.2 RDD 常用算子之 	<p>时间：1.21 (9:00-18:00)</p> <ol style="list-style-type: none"> 配置 Spark IDEA 开发环境 <ol style="list-style-type: none"> 5.1 搭建 Spark 开发环境 Spark SQL 应用 <ol style="list-style-type: none"> 6.1 Spark SQL 简介 6.2 Spark SQL 配置 6.3 从结构化数据文件创建 DataFrame 6.4 从外部数据库创建 DataFrame 6.5 从 RDD 创建 DataFrame 6.6 读取 Hive 表数据创建 DataFrame 6.7 读取学生成绩创建 DataFrame 6.8 常见 DataFrame API 操作 1 6.9 常见的 DataFrame 操作 2 6.10 常见的 DataFrame 操作 3 6.11 通过 DataFrame API 计算学生总分和平均分 6.12 保存 DataFrame 数据 6.13 保存学生成绩分析结果到 Hive 6.14 DataSet 基础操作 6.15 统计商品销量 <p>拓展内容：</p> <ol style="list-style-type: none"> HBase 非关系型数据库 Zookeeper 大数据分布式消息 Flume 数据采集 Kafka 消息系统 Flink 大数据实时处理 商品实时推荐系统 <p>特别内容：</p> <ol style="list-style-type: none"> ① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想

3.2 Hadoop 集群基本信息介绍 3.3 HDFS 常用 Shell 操作 3.4 MapReduce 常用 Shell 操作 3.5 MapReduce 任务管理 3.6 Yarn 资源管理与任务调度 4. MapReduce 开发入门 4.1 MapReduce 框架与设计构思 4.2 MapReduce 开发环境搭建 4.3 MapReduce WordCount 源码分析 4.4 MapReduce API 分析 4.5 MapReduce 统计每天访问次数 4.6 MapReduce 按用户访问次数排序 拓展 5. MapReduce 编程进阶	transformation 算子 (1) 4.3 RDD 常用算子之 transformation 算子 (2) 4.4 RDD 常用算子之键值对 RDD 算子 4.5 RDD 常用算子之 action 算子 4.6 文件读取与存储 4.7 统计用户停留时间最长的基站	
---	--	--

案例实战课程安排

时间：1.22 (9:00-18:00) 航空客户价值分析 1. 案例背景与挖掘目标 2. 数据存储 3. 数据探索与数据清洗 4. 构建航空客户价值分析关键特征 5. 航空客户分群 6. 会员营销策略与建议	时间：1.23 (9:00-18:00) 广电大数据用户画像 1. 项目背景与目标分析 1.1 背景介绍 1.2 目标分析 1.3 系统架构介绍 2. 数据存储 2.1 数据介绍 2.2 数据存储 2.3 开发环境搭建 3. 数据探索与清洗	时间：1.24 (9:00-18:00) 4. 用户画像标签计算 4.1 用户画像简介 4.2 消费内容 4.3 消费等级 4.4 产品名称 4.5 入网程度&业务名称 4.6 地区&语言偏好 5. SVM 预测用户是否挽留 5.1 SVM 介绍 5.2 特征构建 5.3 标签构建 5.4 模型构建与评估 时间：1.25 (16:00-18:00) 职业技术考试 跳转至线上课程安排
---	--	--

专题五 商务数据分析实战 (Excel+Power BI) 课程大纲

技能学习课程安排

时间：1.16 (9:00-18:00) 专题讲座 计算机视觉技术及其应用 Excel 数据分析基础与实战 1.1 认识数据分析	时间：1.17 (9:00-18:00) 6.1 对比分析 6.2 趋势分析 6.3 饼图 6.4 散点图	时间：1.19 (9:00-18:00) 4.1 认识 DAX 语言 4.2 构建日历表 4.3 构建表间关系 4.4 度量值
---	--	--

<p>1.2 认识 Excel 2016</p> <p>2.1 获取文本数据</p> <p>2.2 从数据库获取数据</p> <p>3.1 排序</p> <p>3.2 筛选</p> <p>3.3 分类汇总</p> <p>4.1 认识公式和函数</p> <p>4.2 数组公式</p> <p>4.3 日期和时间函数</p> <p>4.4 数学函数</p> <p>4.5 统计函数</p> <p>4.6 文本函数</p> <p>4.7 逻辑函数</p> <p>5.1 透视表的创建和修改</p> <p>5.2 透视表的操作</p> <p>5.3 透视图的操作</p>	<p>6.5 雷达图</p> <p>7.1 案例背景</p> <p>7.2 数据预处理</p> <p>8 商品销售分析</p> <p>9 库存分析</p> <p>10 用户分析</p> <p>11 分析报告</p> <p>时间：1.18 (9:00-18:00)</p> <p>1.1 认识数据分析</p> <p>1.2 了解可视化工具</p> <p>1.3 认识 power bi</p> <p>2.1 数据来源</p> <p>2.2 获取数据</p> <p>3.1 认识编辑器和 M 语言</p> <p>3.2 数据集成</p> <p>3.3 数据清洗</p> <p>3.4 转换数据</p> <p>3.5 数据泛化</p>	<p>5.1 认识可视化</p> <p>5.2 条形图和柱状图</p> <p>5.3 雷达图和漏斗图</p> <p>时间：1.20 (9:00-18:00)</p> <p>5.4 饼图和环形图</p> <p>5.5 瀑布图和树状图</p> <p>5.6 折线图和散点图</p> <p>5.7 描述性分析</p> <p>5.8 KPI 分析</p> <p>6.1 认识分析报表</p> <p>6.2 制作分析报表</p> <p>7 部署</p> <p>拓展内容</p> <p>Python 编程基础</p> <p>1 认识 Python</p> <p>2 编写 Python 程序</p> <p>3 认识 Python 数据结构</p> <p>4 条件判断及分支语句</p> <p>5 使用 def 定义函数</p> <p>6 认识面向对象</p> <p>7 读取文件数据</p> <p>8 模块和第三方库</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT，让应用开发更高效</p> <p>③ 大模型(ChatGLM2-6B)产业应用漫步与畅想</p>
案例实战课程安排		
<p>时间：1.21 (9:00-18:00)</p> <p>新零售智能销售数据分析(Power BI)</p> <p>1 案例背景</p> <p>2 数据清洗</p> <p>3 数据规约</p> <p>4 数据建模</p> <p>5 销售分析</p> <p>6 库存分析</p> <p>7 用户分析</p> <p>8 部署和发布</p>	<p>时间：1.23 (9:00-18:00)</p> <p>财务分析在纳税评估中的应用(Excel)</p> <p>1 案例背景和分析流程</p> <p>2 数据预处理</p> <p>3 发现疑点</p> <p>4 共同比分析</p> <p>5 增长趋势分析</p> <p>6 财务比率分析</p> <p>7 重点评估区域</p> <p>8 重点评估区域审计</p> <p>9 问题发现和财务报表的调整</p>	<p>时间：1.24 (9:00-18:00)</p> <p>学生校园卡消费行为分析(Excel)</p> <p>1.1 案例背景与目标</p> <p>2.1 预处理：读取数据和异常值</p> <p>2.2 预处理：缺失值</p> <p>2.3 预处理：重复值与合并数据</p> <p>3.1 食堂消费数据分析</p> <p>3.2 学生消费行为分析</p> <p>时间：1.25 (16:00-18:00)</p> <p>职业技术考试</p> <p>跳转至线上课程安排</p>

<p>时间：1.22 (9:00-18:00)</p> <p>餐饮企业综合分析 (Power BI)</p> <p>1 案例背景</p> <p>2 数据预处理</p> <p>2.1 数据预处理 1</p> <p>2.2 数据预处理 2</p> <p>3 数据可视化</p> <p>3.1 数据分析与可视化 1</p> <p>3.2 数据分析与可视化 2</p>	<p>10 小结</p>	
--	--------------	--

专题六 商务数据分析实战 (Python) 课程大纲

<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <p>1 认识 Python</p> <p>2 编写 Python 程序</p> <p>3 认识 Python 数据结构</p> <p>4 条件判断及分支语句</p> <p>5 使用 def 定义函数</p> <p>6 认识面向对象</p> <p>7 读取文件数据</p> <p>8 模块和第三方库</p> <p>Python 数据分析与应用</p> <p>1 Python 数据分析概述</p> <p>2 NumPy 数值计算基础</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 数据分析基础</p> <p>1 Matplotlib 数据可视化基础</p> <p>1.1 掌握绘图基础语法与常用</p> <p>1.2 分析特征间关系</p> <p>1.3 分析特征内部数据分布与分散情况</p>	<p>时间：1.17 (9:00-18:00)</p> <p>2 Pandas 统计分析基础</p> <p>2.1 Pandas 简介</p> <p>2.2 读写不同数据源的数据</p> <p>2.3 数据框与数据框元素</p> <p>2.4 转换与处理时间序列数据</p> <p>2.5 使用分组聚合进行组内计算</p> <p>2.6 创建透视表与交叉表</p> <p>时间：1.18 (9:00-18:00)</p> <p>3 使用 Pandas 进行数据预处理</p> <p>3.1 合并数据</p> <p>3.2 清洗数据</p> <p>3.3 标准化数据</p> <p>3.4 转换数据</p>	<p>时间：1.19 (9:00-18:00)</p> <p>Python 数据分析实训</p> <p>1 iris 数据处理实训</p> <p>1.1 拓展学习资料&Python 环境介绍</p> <p>1.2 读取数据&修改列名称</p> <p>1.3 以 PythonConsole 方式执行代码</p> <p>1.4 缺失值处理</p> <p>1.5 重置索引</p> <p>2 探索 chipotle 数据</p> <p>2.1 数据读取及介绍</p> <p>2.2 分组聚合</p> <p>2.3 数据类型转化</p> <p>2.4 求客单价</p> <p>3 探索 Apple 公司股价数据</p> <p>3.1 数据读取及介绍</p> <p>3.2 找到最后一个交易日</p> <p>3.3 日期探索及可视化</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT，让应用开发更高效</p> <p>③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想</p>
<p>案例实战 课程安排</p>		
<p>时间：1.20 (9:00-18:00)</p> <p>商务数据分析概述</p> <p>1. 商务数据分析基本概念</p> <p>2 商务数据分析应用场景</p> <p>3 商务数据分析主要工作及流程</p>	<p>时间：1.22 (9:00-18:00)</p> <p>聚类分析-对某航空公司客户分群</p> <p>1 背景与目标</p> <p>2 目标分析</p>	<p>时间：1.24 (9:00-18:00)</p> <p>关联规则-购物篮商品推荐</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索</p>

<p>4 商务数据分析方法与工具</p> <p>5 商务数据分析产业需求</p> <p>牛刀小试-T公司运营数据分析与业务决策</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索</p> <p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p> <p>时间：1.21 (9:00-18:00)</p> <p>回归分析-某市财政收入预测</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索</p> <p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p>	<p>3 数据探索</p> <p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p> <p>时间：1.23 (9:00-18:00)</p> <p>分类分析-运营商用户流失判别</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索</p> <p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p>	<p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p> <p>时间：1.25 (9:00-18:00)</p> <p>协同过滤-新闻网站用户个性化推荐</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索</p> <p>4 数据处理</p> <p>5 统计分析</p> <p>6 决策建议</p> <p>时间：1.26 (16:00-18:00)</p> <p>职业技术考试</p> <p>跳转至线上课程安排</p>
---	--	--

专题七 计算机视觉应用实战 (PyTorch) 课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <p>1 认识 Python</p> <p>2 编写 Python 程序</p> <p>3 认识 Python 数据结构</p> <p>4 条件判断及分支语句</p> <p>5 使用 def 定义函数</p> <p>6 认识面向对象</p> <p>7 读取文件数据</p> <p>8 模块和第三方库</p> <p>Python 数据分析与应用</p> <p>1 Python 数据分析概述</p> <p>2 NumPy 数值计算基础</p> <p>3 Matplotlib 数据可视化基础</p> <p>4 Pandas 统计分析基础</p> <p>5 使用 Pandas 进行数据预处理</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 机器学习实战</p>	<p>时间：1.17 (9:00-18:00)</p> <p>5 神经网络</p> <p>5.1 单个神经元介绍</p> <p>5.2 经典网络结构介绍</p> <p>5.3 神经网络工作流程演示</p> <p>5.4 如何修正网络参数-梯度下降法</p> <p>5.5 网络工作原理推导</p> <p>5.6 网络搭建准备</p> <p>5.7 样本从输入层到隐层传输的 Python 实现</p> <p>5.8 网络输出的 Python 实现</p> <p>5.9 单样本网络训练的 Python 实现</p> <p>5.10 全样本网络训练的 Python 实现</p> <p>5.11 网络性能评价</p> <p>5.12 调用 sklearn 实现神经网络</p> <p>时间：1.18 (9:00-18:00)</p>	<p>时间：1.19 (9:00-18:00)</p> <p>PyTorch 深度学习原理与实现</p> <p>1 引言</p> <p>2 卷积神经网络 CNN</p> <p>2.1 浅层神经网络的局限</p> <p>2.2 卷积操作</p> <p>2.3 卷积操作的优势</p> <p>2.4 池化及全连接</p> <p>2.5 高维输入及多 filter 卷积</p> <p>2.6 实现卷积操作</p> <p>2.7 将卷积结果可视化</p> <p>2.8 实现池化操作</p> <p>时间：1.20 (9:00-18:00)</p> <p>3 循环神经网络 RNN</p> <p>3.1 循环神经网络简介</p> <p>3.2 循环神经网络的常见结构</p> <p>4 长短时记忆网络 LSTM</p> <p>4.1 LSTM 的三个门</p> <p>4.2 LSTM 三个门的计算示例</p> <p>4.3 实现 LSTM 操作</p>

<p>1 机器学习绪论</p> <p>1.1 引言</p> <p>1.2 基本术语</p> <p>1.3 假设空间&归纳偏好</p> <p>2 模型评估与选择</p> <p>2.1 经验误差与过拟合</p> <p>2.2 评估方法</p> <p>2.3 性能度量</p> <p>2.4 性能度量 Python 实现</p> <p>3 回归分析</p> <p>3.1 线性回归基本形式</p> <p>3.2 线性回归模型的 Python 实现</p> <p>4 决策树</p> <p>4.1 从女生相亲到决策树</p> <p>4.2 明天适合打球吗</p> <p>4.3 决策树拆分属性选择</p> <p>4.4 决策树算法家族</p> <p>4.5 泰坦尼克号生还者预测—数据预处理</p> <p>4.6 泰坦尼克号生还者预测—模型构建与预测</p>	<p>PyTorch 框架基础实践</p> <p>1 PyTorch 简介</p> <p>2 张量操作</p> <p>2.1 创建张量</p> <p>2.2 张量与数组相互转化</p> <p>3 构建一个线性模型</p> <p>3.1 任务描述：构建线性模型</p> <p>3.2 读取数据</p> <p>3.3 构建初始模型及损失函数</p> <p>3.4 test-构建优化器</p> <p>3.4 构建优化器</p> <p>3.5 最小化方差（训练）</p> <p>3.6 执行多轮训练</p> <p>3.7 训练过程可视化</p> <p>4 识别手写数字</p> <p>4.1 案例目标与流程</p> <p>4.2 加载数据</p> <p>4.3 加工数据</p> <p>4.4 模型结构介绍</p> <p>4.5 构建模型</p> <p>4.6 模型配置</p> <p>4.7 模型训练</p> <p>4.8 执行多轮训练</p> <p>4.9 模型性能评估</p> <p>4.10 保存模型</p> <p>4.11 加载模型</p> <p>4.12 模型应用</p>	<p>4.4 LSTM 返回值解读</p> <p>5 利用 LSTM 实现手写数字识别</p> <p>5.1 加载数据</p> <p>5.2 数据加工</p> <p>5.3 搭建循环神经网络</p> <p>5.4 模型配置</p> <p>5.5 模型训练</p> <p>5.6 模型性能验证</p> <p>时间：1.21（9:00-18:00）</p> <p>计算机视觉实战</p> <p>1 概述</p> <p>2.1 图像基础</p> <p>2.2 读写图像</p> <p>3.1 几何变换-图像平移和旋转</p> <p>3.2 几何变换-最近邻插值</p> <p>3.3 几何变换-其他插值方法介绍</p> <p>4.1 灰度处理-线性变换</p> <p>4.2 灰度处理-非线性变换</p> <p>4.3 灰度处理-直方图均衡化</p> <p>4.4 图像二值化</p> <p>5.1 图像平滑</p> <p>5.2 图像锐化-Sobel 算子</p> <p>5.3 图像锐化-其他算法</p> <p>6. 练习</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT, 让应用开发更高效</p> <p>③ 大模型 (ChatGLM2-6B) 产业应用漫步与畅想</p>
案例实战课程安排		
<p>时间：1.22（9:00-18:00）</p> <p>脑 PET 图像分析与疾病预测</p> <p>1 背景与目标</p> <p>2 数据处理</p> <p>2.1 图像读取及尺寸调整</p> <p>2.2 图像增强之翻转操作</p> <p>2.3 图像增强之旋转缩放</p> <p>2.4 获取所有照片路径</p> <p>2.5 批量获取照片数据</p> <p>2.6 将数据处理过程封装成函数</p> <p>3 模型构建</p>	<p>时间：1.23（9:00-18:00）</p> <p>基于 FaceNet 的人脸智能识别</p> <p>1 背景与目标</p> <p>1.1 背景与目标</p> <p>1.2 目标分析</p> <p>1.3 开发环境和工程结构介绍</p> <p>2 人脸检测</p> <p>2.1 MTCNN 人脸检测介绍</p> <p>2.2 执行人脸检测操作</p> <p>3 人脸对齐</p> <p>3.1 执行人脸对齐操作</p>	<p>时间：1.24（9:00-18:00）</p> <p>基于 YOLOX 的农田害虫图像检测与识别</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>2.1 目标检测任务的难点与挑战</p> <p>2.2 目标检测方法发展历程</p> <p>2.3 经典二阶段&一阶段算法</p> <p>3 YOLO 目标检测</p> <p>3.1 将目标检测转化为回归任务</p> <p>3.2 YOLOv1 中训练样本标签设计</p>

<ul style="list-style-type: none"> 3.1 定义卷积&池化层 3.2 定义全连接层 3.3 定义网络计算过程 3.4 数据集维度顺序调整及类型转化 3.5 数据集分批及打乱操作 3.6 模型构建及配置 3.7 模型训练 4 模型性能评估及应用 4.1 模型性能评估及保存 4.2 模型应用 	<ul style="list-style-type: none"> 3.2 人脸检测及对齐代码整理 4 人脸特征提取 4.1 FaceNet 溯源-计算机视觉领域的部分大事件 4.2 FaceNet 介绍 4.3 执行人脸特征提取操作 5 人脸识别 5.1 获取后台数据库中的人脸数据 5.2 获取后台数据库人脸数据脚本解读 5.3 完成人脸识别操作 5.4 代码整理与结果可视化 6 小结 	<ul style="list-style-type: none"> 3.3 YOLOv1 网络结构及输出解读 3.4 YOLOv1 损失函数介绍 3.5 YOLOv1 的缺点 3.6 YOLOv3 及 YOLOX 介绍 3.7 目标检测常用数据集与性能评价指标介绍 3.8 项目目标完成步骤介绍 4 数据探索与处理 4.1 配套资源说明 4.2 数据探索 4.3 CopyAndPast 数据增强介绍 4.4 CopyAndPast 数据增强实现 4.5 汇总照片数据 时间：1.25 (9:00-18:00) 5 数据加工 5.1 数据加工介绍 5.2 数据加工实现 6 环境搭建与模型训练 6.1 创建虚拟开发环境 6.2 启动虚拟环境并为其安装依赖库 6.3 安装 YOLOX 6.4 模型训练 7 模型应用 7.1 模型应用 7.2 小结 8 拓展延伸 时间：1.26 (16:00-18:00) 职业技术考试 跳转至线上上班课程安排
---	---	---

专题八 文本分析与挖掘实战 (PyTorch) 课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <ul style="list-style-type: none"> 1 认识 Python 2 编写 Python 程序 3 认识 Python 数据结构 4 条件判断及分支语句 5 使用 def 定义函数 6 认识面向对象 	<p>时间：1.17 (9:00-18:00)</p> <ul style="list-style-type: none"> 4.6 网络搭建准备 4.7 样本从输入层到隐层传输的 Python 实现 4.8 网络输出的 Python 实现 4.9 单样本网络训练的 Python 实现 4.10 全样本网络训练的 Python 	<p>时间：1.19 (9:00-18:00)</p> <p>PyTorch 深度学习原理与实现</p> <ul style="list-style-type: none"> 1 引言 2 卷积神经网络 CNN 2.1 浅层神经网络的局限 2.2 卷积操作 2.3 卷积操作的优势 2.4 池化及全连接

<p>7 读取文件数据</p> <p>8 模块和第三方库</p> <p>Python 数据分析与应用</p> <p>1 Python 数据分析概述</p> <p>2 NumPy 数值计算基础</p> <p>3 Matplotlib 数据可视化基础</p> <p>4 Pandas 统计分析基础</p> <p>5 使用 Pandas 进行数据预处理</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 机器学习实战</p> <p>1 机器学习绪论</p> <p>1.1 引言</p> <p>1.2 基本术语</p> <p>1.3 假设空间&归纳偏好</p> <p>2 模型评估与选择</p> <p>2.1 经验误差与过拟合</p> <p>2.2 评估方法</p> <p>2.3 性能度量</p> <p>2.4 性能度量 Python 实现</p> <p>3 回归分析</p> <p>3.1 线性回归基本形式</p> <p>3.2 线性回归模型的 Python 实现</p> <p>4 人工神经网络</p> <p>4.1 单个神经元介绍</p> <p>4.2 经典网络结构介绍</p> <p>4.3 神经网络工作流程演示</p> <p>4.4 如何修正网络参数-梯度下降法</p> <p>4.5 网络工作原理推导</p>	<p>实现</p> <p>4.11 网络性能评价</p> <p>4.12 调用 sklearn 实现神经网络算法</p> <p>5.1 聚类分析概述</p> <p>5.2 相似性度量</p> <p>5.3 K-Means 聚类分析算法介绍</p> <p>5.4 利用 K-Means 算法对鸮尾花进行聚类</p> <p>5.5 聚类结果的性能度量</p> <p>5.6 调用 sklearn 实现聚类分析</p> <p>6.1 间隔与支持向量</p> <p>6.2 对偶问题</p> <p>6.3 核函数</p> <p>6.4 软间隔与正则化</p> <p>6.5 支持向量机算法的 Python 实现</p> <p>时间：1.18 (9:00-18:00)</p> <p>PyTorch 框架基础实践</p> <p>1 PyTorch 简介</p> <p>2 张量操作</p> <p>2.1 创建张量</p> <p>2.2 张量与数组相互转化</p> <p>3 构建一个线性模型</p> <p>3.1 任务描述：构建一个线性模型</p> <p>3.2 读取数据</p> <p>3.3 构建初始模型及损失函数</p> <p>3.4 test-构建优化器</p> <p>3.4 构建优化器</p> <p>3.5 最小化方差（训练）</p> <p>3.6 执行多轮训练</p> <p>3.7 训练过程可视化</p> <p>4 识别手写数字</p> <p>4.1 案例目标与流程</p> <p>4.2 加载数据</p> <p>4.3 加工数据</p> <p>4.4 模型结构介绍</p> <p>4.5 构建模型</p> <p>4.6 模型配置</p> <p>4.7 模型训练</p> <p>4.8 执行多轮训练</p> <p>4.9 模型性能评估</p>	<p>2.5 高维输入及多 filter 卷积</p> <p>2.6 实现卷积操作</p> <p>2.7 将卷积结果可视化</p> <p>2.8 实现池化操作</p> <p>时间：1.20 (9:00-18:00)</p> <p>3 循环神经网络 RNN</p> <p>3.1 循环神经网络简介</p> <p>3.2 循环神经网络的常见结构</p> <p>4 长短时记忆网络 LSTM</p> <p>4.1 LSTM 的三个门</p> <p>4.2 LSTM 三个门的计算示例</p> <p>4.3 实现 LSTM 操作</p> <p>4.4 LSTM 返回值解读</p> <p>5 利用 LSTM 实现手写数字识别</p> <p>5.1 加载数据</p> <p>5.2 数据加工</p> <p>5.3 搭建循环神经网络</p> <p>5.4 模型配置</p> <p>5.5 模型训练</p> <p>5.6 模型性能验证</p> <p>时间：1.21 (9:00-18:00)</p> <p>自然语言处理实战</p> <p>1 自然语言处理概述</p> <p>1.1 自然语言处理概述</p> <p>2 NLP 的基本流程</p> <p>2.1 NLP 的基本流程</p> <p>2.2 语料字符处理</p> <p>2.3 分词和停用词处理</p> <p>2.4 N 元语法模型</p> <p>2.5 隐马尔可夫概述</p> <p>时间：1.22 (9:00-18:00)</p> <p>2.6 jieba 分词</p> <p>2.7 去停用词</p> <p>3 文本向量化</p> <p>3.2 词袋模型</p> <p>3.3 TF-IDF</p> <p>3.4 Word2Vec 模型</p> <p>3.5 Doc2Vec 模型</p> <p>特别内容</p> <p>1. 金牌助手 ChatGPT, 让应用开发更高效</p>
---	--	--

	4.10 保存模型 4.11 加载模型 4.12 模型应用	
案例实战课程安排		
时间：1.23 (9:00-18:00) 基于 textCNN 的公众健康问句分类 1 背景与目标 1.1 背景与目标 1.2 目标分析 2 数据探索与处理 2.1 数据探索 2.2 剔除无效字符及分词操作 2.3 读取停用词表 2.4 去除停用词 2.5 为词语编号 3 词嵌入 (Word2Vec) 3.1 词嵌入 (Word2Vec) 介绍 3.2 获取目标词向量矩阵 3.3 保存处理好的数据 4 模型构建 (textCNN) 4.1 统一各样本的词语数量 4.2 textCNN 介绍 4.3 构建 textCNN 4.4 建模前数据准备 4.5 执行模型训练 5 模型性能评估 5.1 模型性能评估 5.2 小结	时间：1.24 (9:00-18:00) 新冠疫情期间网民情绪识别 1 背景与目标 2 数据准备 2.1 数据基本介绍 2.2 词嵌入介绍 2.3 进行词向量训练 2.4 构建词向量矩阵 2.5 获取编码后的语料库 2.6 对各样本执行 padding 操作 2.7 拆分数据并将其转为模型所需格式 3 模型训练与性能验证 3.1 Embedding 层介绍 3.2 定义 Embedding 层算子 3.3 定义 LSTM 层算子 3.4 定义全连接层算子 3.5 定义网络计算流程 3.6 模型构建及配置 3.7 模型训练 3.8 模型性能评估 4 小结	时间：1.25 (9:00-18:00) 推荐系统受众性别智能识别 1 背景与目标 1.1 项目目标 1.2 数据介绍 1.3 目标分析 2 数据探索与处理 2.1 数据读取与探索 2.2 获取用户点击流介绍 2.3 统计用户的 id 点击次数 2.4 获取用户点击流 2.5 提取并保存样本标签 2.6 封装数据处理函数 3 词嵌入 3.1 词嵌入准备 3.2 词向量训练 3.3 构建词向量矩阵 3.4 对点击流进行编码 3.5 统一点击流的长度 3.6 保存词向量及点击流数据 4 模型构建与训练 4.1 建模前数据准备 4.2 模型训练 4 模型性能评估与优化 5.1 模型性能评估 5.2 模型优化 时间：1.26 (16:00-18:00) 职业技术考试 跳转至线上上班课程安排

专题九 网络舆情与情感分析 (Python) 课程大纲

技能学习课程安排		
时间：报名成功后即可开始学习 Python 编程基础 1 认识 Python 2 编写 Python 程序 3 认识 Python 数据结构 4 条件判断及分支语句	时间：1.18 (9:00-18:00) 5.1 聚类分析概述 5.2 相似性度量 5.3 K-Means 聚类分析算法介绍 5.4 利用 K-Means 算法对鸢尾花进行聚类	时间：1.21 (9:00-18:00) Python 文本挖掘实战 1 自然语言处理简介 2 开源中文 NLP 系统介绍 3 中文分词介绍 4 机械分词法

<p>5 使用 def 定义函数</p> <p>6 认识面向对象</p> <p>7 读取文件数据</p> <p>8 模块和第三方库</p> <p>Python 数据分析与应用</p> <p>1 Python 数据分析概述</p> <p>2 NumPy 数值计算基础</p> <p>3 Matplotlib 数据可视化基础</p> <p>4 Pandas 统计分析基础</p> <p>5 使用 Pandas 进行数据预处理</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>网络舆情与情感分析概述</p> <p>1. 网络舆情概论</p> <p>2. 网络舆情分析技术</p> <p>3. 网络信息采集技术</p> <p>4. 文本分割技术基础概念</p> <p>5. 文本情感分析技术基本概念</p> <p>6. 情感信息分类</p> <p>7. 情感信息的检索与归纳</p> <p>时间：1.17 (9:00-18:00)</p> <p>Python 机器学习实战</p> <p>1 机器学习绪论</p> <p>1.1 引言</p> <p>1.2 基本术语</p> <p>1.3 假设空间&归纳偏好</p> <p>2 模型评估与选择</p> <p>2.1 经验误差与过拟合</p> <p>2.2 评估方法</p> <p>2.3 性能度量</p> <p>2.4 性能度量 Python 实现</p> <p>3 回归分析</p> <p>3.1 线性回归基本形式</p> <p>3.2 线性回归模型的 Python 实现</p> <p>4 决策树</p> <p>4.1 从女生相亲到决策树</p> <p>4.2 明天适合打球吗</p> <p>4.3 决策树拆分属性选择</p> <p>4.4 决策树算法家族</p> <p>4.5 泰坦尼克号生还者预测—数据预处理</p>	<p>5.5 聚类结果的性能度量</p> <p>5.6 调用 sklearn 实现聚类分析</p> <p>6.1 间隔与支持向量</p> <p>6.2 对偶问题</p> <p>6.3 核函数</p> <p>6.4 软间隔与正则化</p> <p>6.5 支持向量机算法的 Python 实现</p> <p>时间：1.19 (9:00-18:00)</p> <p>Python 网络爬虫实战</p> <p>1 Python 爬虫环境与爬虫简介</p> <p>1.1 认识爬虫</p> <p>1.2 认识反爬虫</p> <p>1.3 配置 Python 爬虫环境</p> <p>2 网页前端基础</p> <p>2.1 认识网络信息传输过程</p> <p>2.2 认识 HTTP</p> <p>3 简单静态网页爬取</p> <p>3.1 认识静态网页</p> <p>3.2 实现 HTTP 请求</p> <p>3.3 解析网页</p> <p>3.4 存储数据</p> <p>时间：1.20 (9:00-18:00)</p> <p>4 认识动态网页</p> <p>4.1 认识动态网页</p> <p>4.2 逆向分析爬取动态网页</p> <p>4.3 使用 Selenium 爬取动态网页</p> <p>5 模拟登录</p> <p>5.1 使用表单登录方法实现模拟登录</p> <p>5.2 使用 Cookie 登录方法实现模拟登录</p> <p>5.3 使用 Selenium 模拟登录</p> <p>拓展</p> <p>6 Scrapy 爬虫</p> <p>6.1 认识 Scrapy</p> <p>6.2 通过 Scrapy 爬取基本页面信息</p> <p>6.3 通过 Scrapy 抓取跳转页面数据</p>	<p>5 机器学习算法分词</p> <p>6 NLP 概率图介绍</p> <p>7 jieba 分词演示</p> <p>8 文本的 one-hot 表达</p> <p>9 tf-idf 权值策略实现</p> <p>10 文本的 TF-IDF 表达</p> <p>11 模型训练与预测</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT，让应用开发更高效</p> <p>③ 大模型(ChatGLM2-6B)产业应用漫步与畅想</p>
--	---	---

4.6 泰坦尼克号生还者预测—模型构建与预测	7 拓展：终端协议及爬取工具介绍	
案例实战课程安排		
<p>时间：1.22 (9:00-18:00)</p> <p>微博疫情评论数据爬虫</p> <p>1 背景与目标</p> <p>2.1 评论结构分析</p> <p>2.2 数据接口分析</p> <p>3.1.1 微博页面接口分析</p> <p>3.1.2 微博 id 获取</p> <p>3.1.3 微博接口数据获取</p> <p>3.1.3 微博接口请求</p> <p>3.1.4 长文本获取</p> <p>3.1.5 单页微博获取</p> <p>3.1.6 单页微博获取函数</p> <p>3.2.1 评论数据接口分析</p> <p>3.2.2 单页评论获取</p> <p>3.2.3 多页评论翻页</p> <p>3.3 评论回复数据爬取</p> <p>3.4 单页微博及评论数据爬取</p> <p>3.5 多线程爬虫</p> <p>4 小结</p>	<p>时间：1.23 (9:00-18:00)</p> <p>基于情感分析的疫情期间网民情绪识别</p> <p>1 背景与目标</p> <p>2.1 情感分析介绍</p> <p>2.2 数据预处理</p> <p>2.3 官方数据删除</p> <p>2.4 BosonNLP 情感词典打分</p> <p>2.5 SnowNLP 情感分析</p> <p>3.1 情感分类-数据预处理</p> <p>3.2 分词和去停用词</p> <p>3.3 TF-IDF 词向量</p> <p>3.4 朴素贝叶斯</p> <p>3.5 朴素贝叶斯情感分类</p> <p>3.6 模型预测和总结</p> <p>时间：1.24 (9:00-18:00)</p> <p>网络问政平台数据爬虫</p> <p>1 背景与目标</p> <p>2.1 网址分析</p> <p>2.2 留言 id 获取</p> <p>2.3 留言详情获取</p> <p>2.4 留言详情回复获取</p> <p>2.5 函数整合</p> <p>2.6 多线程爬取</p>	<p>时间：1.25 (9:00-18:00)</p> <p>智慧政务下的问政舆情分析</p> <p>1 背景与目标分析</p> <p>2.1 样本量分析</p> <p>2.2 重复数据处理</p> <p>2.3 文本预处理</p> <p>2.4 分词和去停用词</p> <p>2.5 词云图分词</p> <p>2.6 词向量获取</p> <p>2.7 SVM 文本分类</p> <p>2.8 模型应用</p> <p>3.1 热点数据预处理</p> <p>3.2 热点数据转化</p> <p>3.3 密度聚类简介</p> <p>3.4 热点聚类</p> <p>3.5 超参数选择</p> <p>3.6 热度排序</p> <p>4 小结</p> <p>时间：1.26 (16:00-18:00)</p> <p>职业技术考试</p> <p>跳转至线上上班课程安排</p>

专题十 TensorFlow 与人工智能实战课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <p>1 认识 Python</p> <p>2 编写 Python 程序</p> <p>3 认识 Python 数据结构</p> <p>4 条件判断及分支语句</p> <p>5 使用 def 定义函数</p> <p>6 认识面向对象</p> <p>7 读取文件数据</p> <p>8 模块和第三方库</p>	<p>时间：1.17 (9:00-18:00)</p> <p>5 人工神经网络</p> <p>5.1 单个神经元介绍</p> <p>5.2 经典网络结构介绍</p> <p>5.3 神经网络工作流程演示</p> <p>5.4 如何修正网络参数-梯度下降法</p> <p>5.5 网络工作原理推导</p> <p>5.6 网络搭建准备</p> <p>5.7 样本从输入层到隐层传输的</p>	<p>时间：1.19 (9:00-18:00)</p> <p>TensorFlow2 深度学习原理与实现</p> <p>1.1 深度神经网络-引言</p> <p>2 卷积神经网络 CNN</p> <p>2.1 浅层神经网络的局限</p> <p>2.2 卷积操作</p> <p>2.3 卷积操作的优势</p> <p>2.4 池化及全连接</p> <p>2.5 高维输入及多 filter 卷积</p>

<p>Python 数据分析与应用</p> <p>1 Python 数据分析概述 2 NumPy 数值计算基础 3 Matplotlib 数据可视化基础 4 Pandas 统计分析基础 5 使用 Pandas 进行数据预处理</p> <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座 计算机视觉技术及其应用</p> <p>Python 机器学习实战</p> <p>1 机器学习绪论</p> <p>1.1 引言 1.2 基本术语 1.3 假设空间&归纳偏好</p> <p>2 模型评估与选择</p> <p>2.1 经验误差与过拟合 2.2 评估方法 2.3 性能度量 2.4 性能度量 Python 实现</p> <p>3 回归分析</p> <p>3.1 线性回归基本形式 3.2 线性回归模型的 Python 实现</p> <p>4 决策树</p> <p>4.1 从女生相亲到决策树 4.2 明天适合打球吗 4.3 决策树拆分属性选择 4.4 决策树算法家族 4.5 泰坦尼克号生还者预测—数据预处理 4.6 泰坦尼克号生还者预测—模型构建与预测</p>	<p>Python 实现</p> <p>5.8 网络输出的 Python 实现 5.9 单样本网络训练的 Python 实现 5.10 全样本网络训练的 Python 实现 5.11 网络性能评价 5.12 调用 sklearn 实现神经网络</p> <p>时间：1.18 (9:00-18:00)</p> <p>TensorFlow2 框架基础实战</p> <p>1 任务 1：构建一个线性模型</p> <p>1.1 tensorflow 介绍 1.2 tensorflow2 常用数据类型和操作 1.3 初始化模型 1.4 构建损失函数 1.5 模型训练及可视化 1.6 使用高阶 API-keras</p> <p>2 任务 2：mnist 手写数字识别</p> <p>2.1 数据读取及探索 2.2 交叉熵 2.3 模型构建及训练 2.4 调用保存好的模型对新样本进行预测</p>	<p>2.6 实现卷积操作 2.7 实现池化操作</p> <p>时间：1.20 (9:00-18:00)</p> <p>3 循环神经网络 RNN</p> <p>3.1 循环神经网络简介 3.2 循环神经网络的常见结构</p> <p>4 长短时记忆网络 LSTM</p> <p>4.1 LSTM 的三个门 4.2 LSTM 三个门的计算示例 4.3 利用 RNN&LSTM 实现 mnist 手写数字识别</p> <p>特别内容：</p> <p>① ChatGPT 教学应用 ② 金牌助手 ChatGPT，让应用开发更高效 ③ 大模型(ChatGLM2-6B)产业应用漫步与畅想</p>
案例实战课程安排		
<p>时间：1.21 (9:00-18:00)</p> <p>脑 PET 图像分析与疾病预测</p> <p>1 背景与目标 2 数据处理</p> <p>2.1 图像读取及尺寸调整 2.2 图像增强之翻转操作 2.3 图像增强之旋转缩放 2.4 获取所有照片路径 2.5 批量获取照片数据 2.6 将数据处理过程封装成函数</p> <p>3 模型构建 (AlexNet)</p>	<p>时间：1.22 (9:00-18:00)</p> <p>基于 textCNN 的公众健康问句分类</p> <p>1 背景与挖掘目标</p> <p>2.1 数据探索 2.2 分词和去停用词 2.3 词序列化和长度对齐</p> <p>3.1 预训练词向量 3.2 词向量矩阵</p> <p>4.1 textCNN 模型搭建 4.2 模型训练</p>	<p>时间：1.23 (9:00-18:00)</p> <p>基于 FaceNet 的人脸智能识别</p> <p>1 背景与目标</p> <p>1.1 背景与目标 1.2 目标分析 1.3 准备工程文件</p> <p>2 人脸检测</p> <p>2.1 MTCNN 介绍及脚本准备 2.2 人脸检测 2.3 标记人脸框 2.4 标记脸部关键点</p>

<ul style="list-style-type: none"> 3.1 AlexNet 介绍 3.2 搭建第一次卷积结构 3.3 完成 AlexNet 搭建 3.4 模型训练 4 模型性能评估 4.1 模型性能评估及预测 4.2 小结 	<ul style="list-style-type: none"> 4.3 模型评估和总结 	<ul style="list-style-type: none"> 3 人脸对齐 3.1 裁剪人脸图像 3.2 人脸对齐 3.3 处理多人脸区域并做函数封装 时间：1.24 (16:00-18:00) 4 人脸特征提取 4.1 FaceNet 基本流程 4.2 FaceNet 模型详解 4.3 获取人脸特征向量 5 人脸识别 5.1 加载后台人脸数据库数据 5.2 人脸识别 6 小结 时间：1.25 (16:00-18:00) 职业技术考试 跳转至线上课程安排
---	---	--

专题十一 Python 金融数据分析与实战课程大纲

技能学习课程安排		
<p>时间：报名成功后即可开始学习</p> <p>Python 编程基础</p> <ul style="list-style-type: none"> 1 认识 Python 2 编写 Python 程序 3 认识 Python 数据结构 4 条件判断及分支语句 5 使用 def 定义函数 6 认识面向对象 7 读取文件数据 8 模块和第三方库 <p>Python 数据分析与应用</p> <ul style="list-style-type: none"> 1 Python 数据分析概述 2 NumPy 数值计算基础 3 Matplotlib 数据可视化基础 <p>时间：1.16 (9:00-18:00)</p> <p>专题讲座</p> <p>计算机视觉技术及其应用</p> <p>Python 数据分析基础</p> <ul style="list-style-type: none"> 1 Pandas 统计分析基础 1.1 Pandas 简介 	<p>时间：1.18 (9:00-18:00)</p> <p>Python 机器学习实战</p> <ul style="list-style-type: none"> 1 机器学习绪论 1.1 引言 1.2 基本术语 1.3 假设空间&归纳偏好 2 模型评估与选择 2.1 经验误差与过拟合 2.2 评估方法 2.3 性能度量 2.4 性能度量 Python 实现 3 回归分析 3.1 线性回归基本形式 3.2 线性回归模型实现 4 决策树 4.1 从女生相亲到决策树 4.2 明天适合打球吗 4.3 决策树拆分属性选择 4.4 决策树算法家族 4.5 泰坦尼克号生还者预测—数据预处理 	<p>时间：1.20 (9:00-18:00)</p> <ul style="list-style-type: none"> 6 支持向量机 6.1 间隔与支持向量 6.2 对偶问题 6.3 核函数 6.4 软间隔与正则化 6.5 支持向量机算法的 Python 实现 7 聚类算法 7.1 聚类分析概述 7.2 相似性度量 7.3 K-Means 聚类分析算法介绍 7.4 利用 K-Means 算法对鸢尾花进行聚类 7.5 聚类结果的性能度量 7.6 调用 sklearn 实现聚类分析 时间：1.21 (9:00-18:00) 1 iris 数据处理实训 1.1 拓展学习资料&Python 环境介绍

<p>1.2 读写不同数据源的数据</p> <p>1.3 数据框与数据框元素</p> <p>1.4 转换与处理时间序列数据</p> <p>1.5 使用分组聚合进行组内计算</p> <p>1.6 创建透视表与交叉表</p> <p>时间：1.17 (9:00-18:00)</p> <p>2 使用 Pandas 进行数据预处理</p> <p>2.1 合并数据</p> <p>2.2 清洗数据</p> <p>2.3 标准化数据</p> <p>2.4 转换数据</p>	<p>4.6 泰坦尼克号生还者预测—模型构建与预测</p> <p>时间：1.19 (9:00-18:00)</p> <p>5 神经网络</p> <p>5.1 单个神经元介绍</p> <p>5.2 经典网络结构介绍</p> <p>5.3 神经网络工作流程演示</p> <p>5.4 修正网络参数-梯度下降法</p> <p>5.5 网络工作原理推导</p> <p>5.6 网络搭建准备</p> <p>5.7 样本从输入层到隐层传输的 Python 实现</p> <p>5.8 网络输出的 Python 实现</p> <p>5.9 单样本网络训练实现</p> <p>5.10 全样本网络训练实现</p> <p>5.11 网络性能评价</p> <p>5.12 调用 sklearn 实现神经网络</p>	<p>1.2 读取数据&修改列名称</p> <p>1.3 以 PythonConsole 方式执行代码</p> <p>1.4 缺失值处理</p> <p>1.5 重置索引</p> <p>2 探索 chipotle 数据</p> <p>2.1 数据读取及介绍</p> <p>2.2 分组聚合</p> <p>2.3 数据类型转化</p> <p>2.4 求客单价</p> <p>3 探索 Apple 公司股价数据</p> <p>3.1 数据读取及介绍</p> <p>3.2 找到最后一个交易日</p> <p>3.3 日期探索及可视化</p> <p>特别内容：</p> <p>① ChatGPT 教学应用</p> <p>② 金牌助手 ChatGPT，让应用开发更高效</p> <p>③ 大模型(ChatGLM2-6B)产业应用漫步与畅想</p>
案例实战课程安排		
<p>时间：1.22 (9:00-18:00)</p> <p>上市公司综合评价</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据准备与处理</p> <p>4 基于总体规模与投资效率指标的综合评价</p> <p>5 基于成长与价值指标的综合评价</p> <p>6 小结</p>	<p>时间：1.23 (9:00-18:00)</p> <p>信用卡高风险客户分析</p> <p>1 背景与目标</p> <p>2 目标分析</p> <p>3 数据探索与处理</p> <p>4 指标构建</p> <p>5 聚类分析</p> <p>6 结果分析</p>	<p>时间：1.24 (9:00-18:00)</p> <p>金融客户产品购买预测</p> <p>1 背景与目标</p> <p>2 数据解读</p> <p>3 目标分析</p> <p>4 数据与处理</p> <p>5 特征构建</p> <p>6 模型训练与评估</p> <p>7 模型优化</p> <p>时间：1.25 (16:00-18:00)</p> <p>职业技术考试</p> <p>跳转至线上课程安排</p>

附件三 师资介绍

冯国灿，博士，中山大学数学学院教授，博士生导师。泰迪杯数据挖掘挑战赛组委会委员，中国工业与应用数学会常务理事，广东省工业与应用数学学会理事长，2000-2002 英国格莱莫根大学数字图像实验室和布拉德福大学数字媒体实验室做博士后研究员。主要从事模式识别、计算机视觉研究，参加主持包括国家自然科学基金等科学基金 20 多项，发表学术论

文 100 余篇,入选 2014-2019 爱思唯尔计算机科学中国高被引学者排行榜。

姜春茂, 博士(后), 福建理工大学计算机学院教授, 硕士生导师。主要研究方向: 智能信息处理与决策, 多粒度云计算, 汉俄自然语言处理等。近年来, 主持和参与国家自然科学基金、黑龙江省自然科学基金、国家社科基金等 10 余项。在 Information sciences、Cognitive computation、Knowledge-based Systems、Applied Intelligence、电子学报等重要期刊发表论文 80 余篇, 其中 SCI 检索 30 余篇。主持教学改革项目 8 项, 包括省级重点, 省级一般等。获得黑龙江省自然科学技术学术成果二等奖, 黑龙江省高校科技三等奖。

樊老师, 大数据开发工程师, Hortonworks 授权 Apache Hadoop 开发者认证培训讲师, Hadoop、Mahout 技术实践者和研究者; 对 Hadoop 的 MapReduce 编程模型有深刻理解, 同时对 Mahout 技术有较深认识, 对 Mahout 源码有深入研究, 擅于 Mahout 中数据挖掘的 K 均值聚类算法、贝叶斯分类算法、FP 树关联规则算法的应用; 主编《Mahout 算法解析与案例实战》、《Hadoop 数据分析与挖掘实战》、《Hadoop 与大数据挖掘》等图书专著; 具有电信行业和银行业的项目经验和行业知识, 主持中国电科院电力大数据平台、电能量数据挖掘与智能分析、客户服务智能分析系统等项目。

刘名军, 广东泰迪智能科技股份有限公司技术副总经理, 中山大学计算机科学与技术专业, 高级信息系统项目管理师。专注于 hadoop 及相关子系统 (hbase/pig/hive), Cassandra, memcached 等云计算相关产品的研究; 具有丰富的计算机软件、数据挖掘算法及应用的经验。曾主持电力行业统计分析决策系统、电力大数据平台、客户停电数据综合分析系统、异常用电智能诊断系统、贵阳民政部数据铁笼项目等多个百万级的项目管理和研发工作, 具有丰富的项目管理和研发工作经验, 给电力、民政、政府等行业进行数据挖掘应用培训十余场。

张敏, 广东泰迪智能科技股份有限公司培训总监, 高级信息系统项目管理师。具有丰富的大数据挖掘、人工智能教学和开发经验, 曾为南方电网、国家电网、格力电器、珠江数码等多个大型企业提供项目研发与维护服务。参编数据挖掘与人工智能类教材 10 余本, 作为主讲老师参与国内高校和企业关于数据挖掘、人工智能相关培训班百余场。

周东平, 广东泰迪智能科技股份有限公司产品总监, 高级数据分析师, 项目经理。从事数据行业多年, 熟悉大数据、人工智能相关项目开发流程; 具有丰富的大数据、人工智能产品与应用设计经验, 对于大数据、人工智能教学具有深入研究。精通 Python 语言, 策划主持

编写 Python 方向大数据与人工智能图书 10 余本，包括《Python 数据分析与应用》、《Python3 智能数据分析快速入门》、《大数据数学基础（Python 语言描述）》等。在职期间参与项目涵盖电力、税务、金融、新闻、建筑等方向，具有丰富的行业经历。

律波，高级数据分析工程师，应用统计学硕士，有较强的统计学、数学、数据挖掘理论功底；精通 R、Python、Power BI、Excel 等数据挖掘分析工具，具有丰富的培训和项目经验，擅长从数据中发掘规律，对数据具有较高的敏感度，逻辑思维能力强，擅长数据可视化，机器学习、深度学习等算法原理的实现，如神经网络、SVM、决策树、贝叶斯等；负责“珠江数码大数据营销推荐应用”项目，完成标签库的构建及产品推荐模型；负责“京东商城产品评论情感分析”项目，完成了评论数据情感评价模型、LDA 主题模型的构建；通过项目案例的转换；负责多个本科类院校数据分析软件培训和毕业生数据分析培训，先后负责广西科技大学、闽江学院、广东石油化工、韩山师范学院、广西师范大学等数据分析软件培训及实训等。多次负责“泰迪杯”数据挖掘大赛题目的构思和实现、赛前培训。大数据专业系列图书编写委员会成员，负责《R 语言与数据挖掘》、《Python 实训案例》、《Excel 可视化案例》等书籍编写工作。

陈四德，广东泰迪智能科技股份有限公司高级数据分析师，统计学专业，对数据统计分析和数据挖掘领域均有较强的理解和理论基础；有造价行业、游戏行业背景和丰富的项目经验，精通行业内的各种指标分析，擅于从多维度分析数据，逻辑性强；擅长 Python、R 语言、MySQL 数据库等工具，能熟练对数据进行数据处理和分析，掌握常用的数据挖掘算法如分类、聚类等，以及深度学习 TensorFlow 的使用。负责“网站会员流失预测”项目，完成数据处理，模型构建；负责“平台 BI 埋点数据入库及数据分析”项目，完成数据盘点、数据指标整理和把控；负责“游戏数据分析”项目，完成产出游戏生态日报、客户价值分群结果、用户流失的预警、用户画像指标的完善和维护，项目经验丰富。负责过西安交大城市学院、福建农林大学、国培师资培训、韩山师范学院数据分析就业班、湖南科技职业技术学院、武汉科技大学、广东机电职业技术学院国培、柳州城市职业技术学院第一届大数据职业技能竞赛指导、吉林大学珠海学院等培训项目，授课经验丰富。负责过“泰迪杯”数据挖掘挑战赛出题及赛题指导。大数据专业系列图书编写委员会成员，负责《Keras 与深度学习实战》、《Python 中文自然语言处理基础与实战》、《深度学习与计算机视觉实战》等书籍编写工作。

焦正升，广东泰迪智能科技股份有限公司资深项目研发工程师、高级信息系统项目经理、高级软件开发工程师，拥有 7 年相关从业经验；致力于信息技术的应用与传播，信息系统产业的发展。精通 JAVA 编程语言，熟悉 Spring Boot、Spring Cloud 等主流开发框架、MySQL 数据库、VUE 数据驱动渐进式框架等主流技术。参与《Hadoop 与大数据挖掘》、《Hadoop 大数据分析与应用》等图书的编写。拥有电力、电子政务、轻工环保、交通运输等多项领域的管理研发经验，项目团队为北京市信访办研发的“大数据助力智慧信访”系统获得第七届金铃奖-公共服务类“智能决策奖”。

郑素铃，从事大数据项目研发工作，对 Hadoop 大数据技术有较深的研究，熟练掌握 Hadoop 环境部署和 Hadoop 核心计算框架 MapReduce 的原理和应用。掌握 Spark 原理及编程，熟练使用 Spark 的图计算 Graphx 和算法库 MLlib。对非结构化数据库 HBase 以及结构化数据库 Hive 有深刻的了解。掌握数据挖掘和机器学习的常用算法，熟悉数据挖掘流程，具备项目开发经验，如“数睿思网站用户画像研究”和“法律服务智能推荐系统”项目，在推荐系统方面比较有研究。先后参与了《Hadoop 大数据开发基础》、《Spark 大数据技术与应用》等图书编写工作。

吴嘉泳，一线大数据研发工程师。对 Hadoop + Spark 生态体系有深入研究，熟练掌握 Hadoop 环境部署和 Hadoop 核心计算框架 MapReduce 的原理和应用。掌握 Spark 原理及编程，熟练使用 Spark 的算法库 MLlib。对非结构化数据库 HBase 以及结构化数据库 Hive 有深刻的了解。深度参与某电网公司内部客服优化系统开发，主要负责对系统中的海量文本数据进行处理和挖掘，利用 Spark + Hive 和相应组件实现潜在规律地挖掘。参与编写 1+X 相关图书编写工作，参与《Hadoop 大数据开发基础》、《Spark 大数据技术与应用》等图书编写。先后跟进负责第三期全国高校大数据与人工智能双师型骨干教师研修班和部分项目案例资源开发工作。

叶丽凡，广东泰迪智能科技股份有限公司高级大数据开发工程师，对 Hadoop 大数据技术有深入理解，熟悉 HDFS 分布式文件系统存储结构，熟练掌握 Hadoop 环境部署和 MapReduce 计算框架编程。对 HBase、Hive 数据库有深刻了解。掌握 Spark 原理及编程，熟悉 Spark 底层运行机制，并熟练使用 Spark SQL 即时查询框架和 Spark MLlib 算法库。深度参与过华南某广电公司大数据营销推荐系统开发，利用 Hadoop + Spark + Hive 为其中的 400 多万用户生成用户画像。参与编写《Hadoop 与大数据挖掘》、《Spark 大数据技术与应用》等图书。

跟进负责全国高校大数据与人工智能双师型骨干师资研修班、韩山师范学院等高校的大数据培训课程。

温鼎，从事大数据研发工作，对 Hadoop 生态圈相关组件技术有深刻的认识，掌握 Hadoop 核心框架原理。熟练掌握 Hadoop、Hive、Flume 等大数据核心组件的环境搭建和使用。同时对 Spark 核心 API core 和 SQL & DataFrame 的使用有深刻的认识，掌握 MLlib 算法库的应用。掌握数据挖掘和机器学习十大算法的应用。对数据有较强的敏感度，掌握常用机器学习算法原理。先后参与湖南商务职业技术学院 Hadoop 模块考题开发，深圳职业计算学院 PySpark 大数据课程资源开发，广东技术师范大学 Hadoop 大数据培训。

陈晓枫，广东泰迪智能科技股份有限公司大数据研发工程师、大数据讲师。对 Hadoop 生态圈相关组件技术有一定的理解，掌握 Hadoop 核心框架原理。熟悉 MapReduce 编程，了解实时框架 HBase、Spark Streaming 和 Flink。参与《Hadoop 大数据应用（第 2 版）》、《Spark 大数据技术与应用（第 2 版）》、《大数据项目实战》等图书的编写和修改。曾参与多个院校教材资源开发工作。曾参与韩山师范学院 3+1 班等大数据培训课程。