

# 全国高校大数据与人工智能师资研修班 邀请函

(2022 年第四期)

---

**四大专题：**数据采集与处理实战（Python）、大数据分析 with 机器学习实战、  
自然语言处理实战（Python）、计算机视觉处理实战（Python）

**主办单位：**泰迪杯数据挖掘挑战赛组委会

**承办单位：**广东泰迪智能科技股份有限公司

**协办单位：**人民邮电出版社有限公司

北京泰迪云智信息技术研究院

互联网、大数据、云计算、人工智能等现代信息技术深刻改变着人类的生产、生活、学习乃至思维方式，深刻展示了世界发展的前景。目前各院校的大数据和人工智能专业教师匮乏、相关落地动手实战应用能力欠缺、授课过程中相关行业实战案例项目缺失等，为加快建设大数据、人工智能相关专业教师队伍，推动各院校建立人才培训和评价体系，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 50 余场，参训教师近 6000 人次。2022 年第四期全国高校大数据与人工智能师资研修班将开设四大专题方向，本期研修班以线上云课堂形式举办，现将有关安排通知如下。



## 一、课程特色

- 1、本研修班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，真正的让学员把所学内容和工作实际有效结合、更好地进行教育教学工作。
- 2、核心课程部分由讲师手把手一起进行实操演练，在具体应用场景中全面掌握相关技能，助力实训教学工作、实际动手的能力。视频制作精良，讲师真人出镜，全面解析专业必备技能，为相关课程开设和备课、应对科研和项目开发打下坚实基础。
- 3、课程设有答疑交流讨论群，培训期间助教全程辅助教学，每天提供 10 小时的实时在线答疑辅导，并进行答疑文档汇总，帮助学员更好地总结学习。
- 4、本课程配套有基础知识内容，即使零基础学员快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、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！视频内容支持六个月内免费回看，以便复习和参考。
- 6、全面实践大数据/人工智能项目流程，包括数据采集、数据存储管理、数据探索、数据处理、特征工程、数据建模等课程，提供知识讲解，助力夯实理论基础，掌握核心技术。
- 7、参加线上课程学习的学员，如后续本人参加线下课程继续深造，持线上缴费凭证可享受一次免费学习机会（仅限参加同一培训专题方向的线下课程）。

## 二、课程安排

|                                                                                                                                                                                                   |                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| 专题一 数据采集与处理实战（Python）                                                                                                                                                                             | 学习时间：05 月 20 日-05 月 30 日，共计 88 学时 |
| 证书颁发：高级 Python 技术应用工程师职业资格证书                                                                                                                                                                      | 费用：1980 元（报名费、学习费、资料费、证书费等）       |
| 课程模块：Python 编程基础、Python 数据分析与应用、Pandas 数据分析基础、Python 数据可视化、Python 网络爬虫实战、数据采集与处理实战：《流浪地球》豆瓣影评采集、数据采集与处理实战：农产品信息采集与分析、数据采集与处理实战：泰迪内推平台招聘信息采集与分析、拓展自学篇：网站图像素材采集实战。<br>详见附件一 数据采集与处理实战（Python）课程大纲   |                                   |
| 专题二 大数据分析与机器学习实战（Python）                                                                                                                                                                          | 学习时间：05 月 21 日-05 月 31 日，共计 88 学时 |
| 证书颁发：高级机器学习工程师职业资格证书                                                                                                                                                                              | 费用：1980 元（报名费、学习费、资料费、证书费等）       |
| 课程模块：Python 编程基础、Python 数据分析与应用、Pandas 数据分析基础、Python 数据可视化、Python 机器学习实战、实战案例：学生校园消费行为分析、实战案例：百货商场用户画像描绘与价值分析、实战案例：O2O 优惠券使用预测、实战案例：泰迪内推平台信息精准推荐应用（基于泰迪建模平台实现）。<br>详见附件二 大数据分析与机器学习实战（Python）课程大纲 |                                   |





|                                                                                                                                                                                                  |                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| 证书颁发：高级大数据技术应用职业技术证书                                                                                                                                                                             | 费用：2480 元（报名费、学习费、资料费、证书费等）       |
| 课程模块： Python 编程基础、Python 数据分析与应用、 Python 机器学习算法原理与实现、TensorFlow2 实战、TensorFlow2 深度学习原理与实现、自然语言处理实战、实战案例：公众健康问句分类、实战案例：搭建一个属于自己的聊天机器人、综合实战：基于深度学习的推荐系统受众性别预测。<br>详见附件三 自然语言处理实战（Python）课程大纲       |                                   |
| 专题四 计算机视觉处理实战（Python）                                                                                                                                                                            | 学习时间：05 月 20 日-05 月 31 日，共计 96 学时 |
| 证书颁发：高级大数据分析师职业技术证书                                                                                                                                                                              | 费用：2480 元（报名费、学习费、资料费、证书费等）       |
| 课程模块： Python 编程基础、Python 数据分析与应用、Python 机器学习算法原理与实现、TensorFlow2 实战、TensorFlow 2 深度学习原理与实现、计算机视觉实战、图像处理实战：脑 PET 图像分析和疾病预测、计算机视觉实战：基于深度学习的肝脏肿瘤分割、实战案例：基于 GAN 的图像风格转换<br>详见附件四 人工智能实战（计算机视觉方向） 课程大纲 |                                   |

### 三、师资介绍

**方海涛**，中国科学院数学与系统科学研究院研究员，博士生导师，《控制理论与应用》杂志副主编，泰迪杯数据挖掘挑战赛专家组成员。主要研究方向：系统估计、优化与控制等。

**张敏**，广东泰迪智能科技股份有限公司、培训总监，从事用户数据分析和数据挖掘工作六年，具有丰富的大数据挖掘理论及实践培训经验，对数据具有较高的敏感度，根据数据对其进行全面的统计分析。精通 Python、R 语言、MATLAB



等多种数据挖掘工具。擅长市场发展情况监控、精确营销方面的数据挖掘工作。有为南方电网、珠江数码等大型企业长期提供实施服务的经验，主导了电子商务网站用户行为分析及网页智能推荐服务、中医证型关联规则挖掘、电信业务话单量预测、航空公司客户价值分析等多个项目。2017 年“泰迪杯数据挖掘挑战赛教练员培训”主讲讲师，2018 年广东省 Python 与深度学习技术师资培训班主讲讲师，2018 年第一/三/五期全国高校大数据核心技术与应用师资研修班主讲讲师、2019 年第一/二/三期全国高校大数据与人工智能师资研修班主讲讲师，2019 年国家电网大数据竞赛河北、湖南省、甘肃省电力系统培训班主讲讲师，先后负责过西安理工大学、广东工业大学、广西师范学院、广西科技大学、闽江学院、广东石油化工学院、上海健康医学院等高校实训课程及德生科技等企业内训和数据挖掘就业班的课程。组织、参与编写图书《Python 编程基础》、《Python 数据分析与应用》、《R 语言编程基础》等。

**律波**，广东泰迪智能科技有限公司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“京东电商产品评论情感分析”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。

## 四、证书颁发

学员经在线培训并考试合格后，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。



## 五、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、本期研修班由广东泰迪智能科技股份有限公司收取费用并开具发票。

3、本期研修班两专题及以上联报者可享受九折优惠。

4、联系方式

联系人：庄丽

电 话：13381084246（微信同号）

邮 箱：[960690465@qq.com](mailto:960690465@qq.com)



扫一扫，添加庄老师微信



泰迪杯数据挖掘挑战赛组委会  
二零二二年四月二十五日



广东泰迪智能科技股份有限公司  
二零二二年四月二十五日





## 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技术证书使用，请真实完整填写。 |                                                                           |     |      |      |          |      |      |
| 姓 名                              | 性 别                                                                       | 职 务 | 毕业院校 | 最高学历 | 手机号      | 电子邮箱 | 专题选择 |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
|                                  |                                                                           |     |      |      |          |      |      |
| 费用支付方式                           | 1、电汇到指定账号。<br>2、扫码支付（报名后联系工作人员索要支付码）。<br>3、付款时请注明“大数据研修班+单位或姓名”字样，方便查账备案。 |     |      |      |          |      |      |
| 账户信息                             | 账户名：广东泰迪智能科技股份有限公司<br>开户行：中国工商银行广州花城支行<br>账户号：3602 0285 0920 1663 221     |     |      |      |          |      |      |
| 备 注                              | 请将报名表发送至邮箱：960690465@qq.com<br>联系人：庄丽 13381084246（微信）                     |     |      |      |          |      |      |

## 附件一 数据采集与处理实战（Python）课程大纲

### 基础篇（报名成功后即可开始学习）

| 时间    | 课程内容                                                                                                                                                                                                                                                                                                                                                                                                                   | 学习平台  |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| 正式培训前 | <b>Python 编程基础</b><br>1 准备工作<br>2 列表操作<br>3 程序流程控制语句<br>4 字符串操作<br>4.1 字符串及其索引&切片<br>4.2 字符串的常见方法<br>4.3 字典的创建及索引<br>4.4 字典常用操作<br>4.5 字典推导式<br>5Python 文件读取操作<br>5.1Python 读取文件<br>5.2 练习 3：统计小说中的单词频次<br>6 函数<br>6.1Python 函数自定义<br>6.2 练习 4：自定义求序列偶数个数的函数<br>7 面向对象与模块<br>7.1Python 方法与函数对比介绍<br>7.2Python 面向对象示例<br>7.3Python 模块使用<br>7.4 第三方库的安装与调用<br>8 注意事项<br>8.1Python 工作路径说明<br>8.2 模块命名及存放路径的注意事项<br>8.3 结语 | 泰迪云课堂 |
| 正式培训前 | <b>Python 数据分析与应用</b><br>1 Python 数据分析概述<br>1.1 认识数据分析<br>1.2 熟悉 Python 数据分析的工具<br>1.3 安装 anaconda 与掌握 Jupyter Notebook 常用功能<br>2 NumPy 数值计算基础                                                                                                                                                                                                                                                                         | 泰迪云课堂 |

|  |                                                                                                                                                                                                                                                                                                                                                                                                         |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | 2.1 掌握 NumPy 数组对象<br>2.1.1 NumPy 简介<br>2.1.2 数组创建及基础属性<br>2.1.3 初识数组的特点<br>2.1.4 创建常用数组<br>2.1.5 数组数据类型<br>2.1.6 生成随机数<br>2.1.7 一维数组的索引<br>2.1.8 逻辑型索引<br>2.1.9 多维数组的索引<br>2.1.10 求解距离矩阵<br>2.1.11 变化数组 shape<br>2.2 掌握 NumPy 矩阵与通用函数<br>2.2.1 NumPy 矩阵介绍<br>2.2.2 NumPy 通用函数介绍<br>2.2.3 通用函数的广播机制<br>2.3 利用 NumPy 进行统计分析<br>2.3.1 NumPy 读写二进制文件<br>2.3.2 NumPy 读写 txt 文件<br>2.3.3 利用 NumPy 对数据进行简单统计分析 |  |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

### 核心课程篇

| 时间                       | 课程内容                                                                                                                                                                                                                       | 学习平台  |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| <b>专题讲座</b>              |                                                                                                                                                                                                                            |       |
| 05 月 20 日<br>18:30-22:00 | 主讲：方海涛<br>主题内容：随机最优控制与人工智能                                                                                                                                                                                                 | 泰迪云课堂 |
| <b>第一课 Pandas 数据分析基础</b> |                                                                                                                                                                                                                            |       |
| 05 月 20 日<br>18:30-22:00 | 1 Pandas 统计分析基础<br>1.1 Pandas 简介<br>1.2 读写不同数据源的数据<br>1.2.1 Pandas 读取文本数据<br>1.2.2 存储数据框<br>1.2.3 Pandas 读取 excel 文件<br>1.2.4 将数据框存储为 excel 文件<br>1.3 数据框与数据框元素<br>1.3.1 构建数据框<br>1.3.2 查看数据框的常用属性<br>1.3.3 按行列顺序访问数据框中的元素 | 泰迪云课堂 |

|                          |                                                                                                                                                                                                                                                       |       |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                          | 1.3.4 按行列名称访问数据框中的元素<br>1.3.5 修改数据框中的元素<br>1.3.6 删除数据框中的元素<br>1.3.7 描述分析数据框中的元素<br>1.4 转换与处理时间序列数据<br>1.4.1 转换成时间类型数据<br>1.4.2 时间类型数据的常用操作<br>1.5 使用分组聚合进行组内计算<br>1.5.1 groupby 分组操作<br>1.5.2 agg 聚合操作<br>1.6 创建透视表与交叉表<br>1.6.1 生成透视表<br>1.6.2 生成交叉表 |       |
|                          | 操作演练                                                                                                                                                                                                                                                  | 个人 PC |
|                          | 在线答疑                                                                                                                                                                                                                                                  | 微信群   |
|                          |                                                                                                                                                                                                                                                       |       |
| 05 月 21 日<br>18:30-22:00 | 2 使用 Pandas 进行数据预处理<br>2.1 合并数据<br>2.1.1 表堆叠<br>2.1.2 主键合并<br>2.1.3 重叠合并<br>2.2 清洗数据<br>2.2.1 检测与处理重复值<br>2.2.2 检测与处理缺失值<br>2.2.3 检测与处理异常值<br>2.3 标准化数据<br>2.4 转换数据<br>2.4.1 哑变量处理<br>2.4.2 离散化连续型数据                                                  | 泰迪云课堂 |
|                          | 操作演练                                                                                                                                                                                                                                                  | 个人 PC |
|                          | 在线答疑                                                                                                                                                                                                                                                  | 微信群   |
|                          |                                                                                                                                                                                                                                                       |       |
| 第二课 Python 数据可视化         |                                                                                                                                                                                                                                                       |       |
| 05 月 22 日<br>18:30-22:00 | 1.1 Matplotlib 绘制流程说明<br>1.2 添加文本和修改绘图风格<br>1.3 rc 参数<br>1.4 散点图<br>1.5 折线图<br>1.6 直方图和条形图<br>1.7 饼图                                                                                                                                                  | 泰迪云课堂 |
|                          |                                                                                                                                                                                                                                                       |       |

|                          |                                                                                                                                                                                                                                                                                                                                              |       |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                          | 1.8 箱线图<br>1.9 人口特征间分布<br>1.10 人口各个特征分布<br>2.1Pyecharts 基础介绍<br>2.2Pyecharts 绘制日历图<br>2.3Pyecharts 绘制漏斗图<br>2.4Pyecharts 绘制仪表盘<br>2.5Pyecharts 绘制水球图<br>2.6Pyecharts 绘制关系图<br>2.7Pyecharts 绘制直角坐标系图表<br>2.8Pyecharts 绘制饼图<br>2.9Pyecharts 绘制雷达图<br>2.10Pyecharts 绘制词云图<br>2.11Pyecharts 绘制柱状图<br>2.12Pyecharts 绘制树形图<br>2.13Pyecharts 绘制地理图表 |       |
|                          | 操作演练                                                                                                                                                                                                                                                                                                                                         | 个人 PC |
|                          | 在线答疑                                                                                                                                                                                                                                                                                                                                         | 微信群   |
| 拓展自学篇                    |                                                                                                                                                                                                                                                                                                                                              |       |
| 自行安排                     | 3.1 seaborn 基础介绍<br>3.2 seaborn 简单绘图<br>3.3 seaborn 绘图风格<br>3.4 调色板 1<br>3.5 调色板 2<br>3.6 关系图<br>3.7 分类图<br>3.8 分布图<br>3.9 回归图<br>3.10 矩阵图<br>3.11 网格图                                                                                                                                                                                       | 泰迪云课堂 |
| 第三课 Python 网络爬虫实战        |                                                                                                                                                                                                                                                                                                                                              |       |
| 05 月 23 日<br>18:30-22:00 | 1Python 爬虫环境与爬虫简介<br>1.1 认识爬虫<br>1.1.1 常见网上冲浪过程<br>1.1.2 爬虫的概念<br>1.1.3 爬虫合法性<br>1.2 认识反爬虫<br>1.3 配置 Python 爬虫环境                                                                                                                                                                                                                             | 泰迪云课堂 |

|                          |                                                                                                                                                                                                                                                                                                                                                                                                                |       |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                          | 2 网页前端基础<br>2.1 认识网络信息传输过程<br>2.1.1 网络传输模型<br>2.1.2 网络信息传输过程<br>2.2 认识 HTTP<br>2.2.1 认识 HTTP<br>2.2.2 熟悉 Cookie<br>3 简单静态网页爬取<br>3.1 认识静态网页<br>3.2 实现 HTTP 请求<br>3.2.1 创建工程<br>3.2.2 生成 HTTP 请求<br>3.2.3 完善 HTTP 请求<br>3.3 解析网页<br>3.3.1 为什么要解析网页<br>3.3.2 初识 Xpath<br>3.3.3 Xpath 相对路径及属性查找<br>3.3.4 使用 BeautifulSoup 解析网页<br>3.3.5 网页解析小结<br>3.4 存储数据<br>3.4.1 认识 chrome 开发者工具<br>3.4.2 任务演练：爬取并存储泰迪科技官网首页数据 |       |
|                          | 操作演练                                                                                                                                                                                                                                                                                                                                                                                                           | 个人 PC |
|                          | 在线答疑                                                                                                                                                                                                                                                                                                                                                                                                           | 微信群   |
|                          |                                                                                                                                                                                                                                                                                                                                                                                                                |       |
| 05 月 24 日<br>18:30-22:00 | 4 认识动态网页<br>4.1 认识动态网页<br>4.2 逆向分析爬取动态网页<br>4.2.1 通过网页源码追踪目标数据文件地址<br>4.2.2 通过开发者工具追踪目标数据文件地址<br>4.2.3 爬取数据并进行保存<br>4.3 使用 Selenium 库爬取动态网页<br>4.3.1 搭建 Selenium 环境<br>4.3.2 利用 Selenium 获取网页数据<br>4.3.3 利用 Selenium 控制点击操作<br>5 模拟登录<br>5.1 使用表单登录方法实现模拟登录<br>5.1.1 模拟登录的过程<br>5.1.2 查找提交入口和表单数据                                                                                                              | 泰迪云课堂 |

|                            |                                                                                                                                                                                                                                                                                                     |       |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                            | 5.1.3 提交表单完成模拟登录<br>5.1.4 使用表单登录的注意事项<br>5.2 使用 Cookie 登录方法实现模拟登录<br>5.3 使用 Selenium 模拟登录                                                                                                                                                                                                           |       |
|                            | 操作演练                                                                                                                                                                                                                                                                                                | 个人 PC |
|                            | 在线答疑                                                                                                                                                                                                                                                                                                | 微信群   |
| 05 月 25 日<br>18:30-22:00   | 6 Scrapy 爬虫<br>6.1 认识 Scrapy<br>6.2 通过 Scrapy 爬取基本页面信息<br>6.2.1 创建项目<br>6.2.2 指定字段及创建 spiders<br>6.2.3 完成 spiders 编写<br>6.2.4 运行程序保存数据<br>6.3 通过 Scrapy 抓取跳转页面数据<br>6.3.1 任务介绍及项目创建<br>6.3.2 获取所有页面的 url<br>6.3.3 获取每个页面的新闻二次页面 url<br>6.3.4 提取各新闻二次页面中的目标数据<br>6.3.5 运行程序保存数据<br>7. 拓展：终端协议及爬取工具介绍 | 泰迪云课堂 |
|                            | 操作演练                                                                                                                                                                                                                                                                                                | 个人 PC |
|                            | 在线答疑                                                                                                                                                                                                                                                                                                | 微信群   |
| 第四课 数据采集与处理实战：《流浪地球》豆瓣影评采集 |                                                                                                                                                                                                                                                                                                     |       |
| 05 月 26 日<br>18:30-22:00   | 1.1 案例背景与挖掘目标<br>2.1 短评数据爬取介绍<br>2.2 安装 selenium 及配置 chrome driver<br>2.3 获取用户名<br>2.4 获取短评正文<br>2.5 设置 cookies<br>2.6 获取用户居住地和入会时间信息<br>2.7 单页数据整理<br>2.8 自定义获取单页数据的函数<br>2.9 判定网页是否已被加载<br>2.10 翻页爬取<br>2.11 代码整理及小结                                                                              | 泰迪云课堂 |
|                            | 操作演练                                                                                                                                                                                                                                                                                                | 个人 PC |
|                            | 在线答疑                                                                                                                                                                                                                                                                                                | 微信群   |
|                            | 3.1 短评正文数据预处理                                                                                                                                                                                                                                                                                       | 泰迪云课堂 |

|                       |                               |       |
|-----------------------|-------------------------------|-------|
| 18:30-22:00           | 3.2 词频统计                      |       |
|                       | 3.3 绘制整体评论数据的词云图              |       |
|                       | 3.4 好评差评词云图绘制及小结              |       |
|                       | 4.1 评分分数分布统计                  |       |
| 05月28日<br>18:30-22:00 | 4.2 短评数量与日期的关系                | 泰迪云课堂 |
|                       | 4.3 短评数量与时刻的关系                |       |
|                       | 4.4 不同评分数量与时间的关系              |       |
|                       | 5.1 评论最多的前十个城市                |       |
|                       | 5.2 评分数量与城市的关系                |       |
|                       | 5.3 总结                        |       |
|                       | 操作演练                          | 个人 PC |
|                       | 在线答疑                          | 微信群   |
|                       | 第五课 数据采集与处理实战：农产品信息采集与分析      |       |
|                       | 1.1 背景与分析目标                   | 泰迪云课堂 |
| 05月28日<br>18:30-22:00 | 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              |       |
| 05月29日<br>18:30-22:00 | 3.4 水稻品种类型数量分析                | 泰迪云课堂 |
|                       | 3.5 主要水稻类型被审定的数量分析            |       |
|                       | 3.6 水稻母本分析                    |       |
|                       | 3.7 主要审定公司分析                  |       |
|                       | 4 总结                          |       |
|                       | 操作演练                          | 个人 PC |
|                       | 在线答疑                          | 微信群   |
|                       | 第六课 数据采集与处理实战：泰迪内推平台招聘信息采集与分析 |       |
|                       | 1 背景与目标                       | 泰迪云课堂 |
| 05月29日<br>18:30-22:00 | 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<br>2.6 爬取及解析二级页面数据<br>2.7 翻页爬取及数据保存<br>3 数据处理<br>3.1 读取已爬取完成的数据<br>3.2 数据预处理操作<br>4 分析与可视化<br>4.1 招聘岗位对学历要求分析<br>4.2 各行业的大数据招聘需求数量分析<br>4.3 不同类型公司的薪资待遇分析<br>4.4 小结 |       |
|                          | 操作演练                                                                                                                                                                               | 个人 PC |
|                          | 在线答疑                                                                                                                                                                               | 微信群   |
| 拓展自学篇：网站图像素材采集实战         |                                                                                                                                                                                    |       |
| 自行安排                     | 1 思路介绍<br>2 单个图片文件爬取<br>3 获取一个页面所有图片网址<br>4 保存所有图片<br>5 翻页爬取更多数据<br>6 PDF 文件规律及问题<br>7 PDF 翻页刷新的网址规律<br>8 获取当前页所有图片网址<br>9 翻页刷新爬取所有图片<br>10 图片拼接成 PDF 文件                           | 泰迪云课堂 |
| 第七课 在线考试                 |                                                                                                                                                                                    |       |
| 05 月 30 日<br>19:00-21:00 | 高级 Python 技术应用工程师技术证书在线考试                                                                                                                                                          | 泰迪云课堂 |

## 附件二 大数据分析 with 机器学习实战（Python）课程大纲

### 基础篇（报名成功后即可开始学习）

| 时间    | 课程内容                            | 学习平台  |
|-------|---------------------------------|-------|
| 正式培训前 | Python 编程基础<br>1 准备工作<br>2 列表操作 | 泰迪云课堂 |

|       |                                                                                                                                                                                                                                                                                                                                                                              |       |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|       | 3 程序流程控制语句<br>4 字符串操作<br>4.1 字符串及其索引&切片<br>4.2 字符串的常见方法<br>4.3 字典的创建及索引<br>4.4 字典常用操作<br>4.5 字典推导式<br>5Python 文件读取操作<br>5.1Python 读取文件<br>5.2 练习 3：统计小说中的单词频次<br>6 函数<br>6.1Python 函数自定义<br>6.2 练习 4：自定义求序列偶数个数的函数<br>7 面向对象与模块<br>7.1Python 方法与函数对比介绍<br>7.2Python 面向对象示例<br>7.3Python 模块使用<br>7.4 第三方库的安装与调用<br>8 注意事项<br>8.1Python 工作路径说明<br>8.2 模块命名及存放路径的注意事项<br>8.3 结语 |       |
| 正式培训前 | <b>Python 数据分析与应用</b><br>1 Python 数据分析概述<br>1.1 认识数据分析<br>1.2 熟悉 Python 数据分析的工具<br>1.3 安装 anaconda 与掌握 Jupyter Notebook 常用功能<br>2 NumPy 数值计算基础<br>2.1 掌握 NumPy 数组对象<br>2.1.1 NumPy 简介<br>2.1.2 数组创建及基础属性<br>2.1.3 初识数组的特点<br>2.1.4 创建常用数组<br>2.1.5 数组数据类型<br>2.1.6 生成随机数<br>2.1.7 一维数组的索引<br>2.1.8 逻辑型索引<br>2.1.9 多维数组的索引                                                    | 泰迪云课堂 |

|  |                                                                                                                                                                                                                              |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | 2.1.10 求解距离矩阵<br>2.1.11 变化数组 shape<br>2.2 掌握 NumPy 矩阵与通用函数<br>2.2.1 NumPy 矩阵介绍<br>2.2.2 NumPy 通用函数介绍<br>2.2.3 通用函数的广播机制<br>2.3 利用 NumPy 进行统计分析<br>2.3.1 NumPy 读写二进制文件<br>2.3.2 NumPy 读写 txt 文件<br>2.3.3 利用 NumPy 对数据进行简单统计分析 |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

### 核心课程篇

| 时间                       | 课程内容                                                                                                                                                                                                                                                                                                                                                                                                                                 | 学习平台  |
|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| <b>专题讲座</b>              |                                                                                                                                                                                                                                                                                                                                                                                                                                      |       |
| 05 月 21 日<br>18:30-22:00 | 主讲：方海涛<br>主题内容：随机最优控制与人工智能                                                                                                                                                                                                                                                                                                                                                                                                           | 泰迪云课堂 |
| <b>第一课 Pandas 数据分析基础</b> |                                                                                                                                                                                                                                                                                                                                                                                                                                      |       |
| 05 月 21 日<br>18:30-22:00 | 1 Pandas 统计分析基础<br>1.1 Pandas 简介<br>1.2 读写不同数据源的数据<br>1.2.1 Pandas 读取文本数据<br>1.2.2 存储数据框<br>1.2.3 Pandas 读取 excel 文件<br>1.2.4 将数据框存储为 excel 文件<br>1.3 数据框与数据框元素<br>1.3.1 构建数据框<br>1.3.2 查看数据框的常用属性<br>1.3.3 按行列顺序访问数据框中的元素<br>1.3.4 按行列名称访问数据框中的元素<br>1.3.5 修改数据框中的元素<br>1.3.6 删除数据框中的元素<br>1.3.7 描述分析数据框中的元素<br>1.4 转换与处理时间序列数据<br>1.4.1 转换成时间类型数据<br>1.4.2 时间类型数据的常用操作<br>1.5 使用分组聚合进行组内计算<br>1.5.1 groupby 分组操作<br>1.5.2 agg 聚合操作 | 泰迪云课堂 |

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/878002015076006101>