

关于举办高等职业教育 2019 年江苏省大数据技术 师资研修班的通知

各有关院校：

近几年，国家相继颁布《关于促进大数据发展的行动纲要》、《大数据产业发展规划 2016-2020》等一系列重要的指导文件，大数据成为国家塑造核心竞争力的战略制高点和经济发展的新引擎。与蓬勃发展的大数据产业相比，大数据专业人才的极度紧缺已成为制约大数据发展的主要瓶颈。截止目前，已有 405 所院校获批“大数据技术与应用”高职专业。大数据人才培养是发展大数据产业的重要支撑，数字时代对院校人才培养提出了新的要求，只有不断探索多学科背景下的大数据专业建设新思路，寻找教学科研工作新方法，提升师资队伍专业素养，才可以为大数据产业发展输送更多优质人才。在此背景下，博雅大数据学院与常州工程职业技术学院展开合作，共同举办“2019 年江苏省大数据技术师资研修班”，开展“大数据的 Python 基础与案例实践”专题培训。

希望通过本次培训帮助大数据及人工智能相关学科教师提升授课技能，夯实基础，进一步提升教师教学能力与科研能力，助力高校师资团队建设，为大数据及人工智能相关专业建设发展提供有力支撑。

一、组织机构

主办单位：北京大数据研究院博雅大数据学院

承办单位：江苏省常州市 常州工程职业技术学院

二、培训安排

- **培训时间：**2019 年 5 月 30 日-6 月 3 日（5 天）
- **培训地点：**江苏省常州市常州工程职业技术学院
- **培训对象：**江苏省各高等院校大数据及人工智能相关学科专业的教研室主任、教学带头人、骨干教师；从事大数据及人工智能相关学科领域研究的科研院所的项目负责人、科研人员、博士生、研究生等。
- **培训课程：**大数据的 Python 基础与案例实践
- **收费标准：**2980 元/人，食宿可统一安排，费用自理。
(最低开班人数 35 人，如因未达到开班人数取消开班，全额退还学费)

三、培训课程及安排

- **培训课程：**大数据的 Python 基础与案例实践
- **课程简介：**

在人工智能时代下，大数据分析与管理、机器学习、深度学习等新技术的提出极大地促进了 Python 语言的发展，2017 年 IEEE Spectrum 发布的编程语言交互式排行榜上，Python 独占榜首。Python 以面向对象、容易学习、跨平台以及由此产生的低维护成本，深受 Python 爱好者喜欢并获得好评。本课程介绍 Python 编程语言的基础内容，通过以下 2 个实践案例，使用 Python 第三方工具

完成数据采集、数据清洗等任务。

实践案例一：某音乐平台热门歌曲评论抓取

xx 音乐是国内主要的音乐在线流媒体服务提供商，用户可以在歌曲页面下发表评论，表达自己的想法。歌曲评论在一定程度上可以反映出歌曲的受欢迎程度，以及歌曲的受众群体。歌曲评论数据包含了丰富的信息可以挖掘。本案例采用案例式教学，使用 Python 第三方库 requests、BeautifulSoup 等从某在线音乐平台抓取 Top 50 歌曲的评论信息，并将数据存入 MySQL 数据库中。

实践案例二：web 服务器的 log 日志清洗

web 服务器的 log 日志由服务器产生，记录网站每天的运行，访问情况。从日志中，我们可以获取网站的页面访问量 (PageView, PV 值)、独立 IP 数、注册用户数。还可根据 web 日志文件，对访问用户进行画像，分析用户的行为特征，实现广告精准投放。本案例使用某网站的 log 日志文件，利用 Python 第三方库 Pandas 进行数据清洗，将原始的 log 文件转换为规整的，有效的数据集。

➤ 课程安排

| 日期 | 时间 | 内容 |
|----------|-------------|---|
| 5 月 30 日 | 14:00-17:00 | 报到 |
| 5 月 31 日 | 9:00-10:00 | 实践案例发布：某音乐平台评论信息抓取和 web 服务器的 log 日志清洗 案例主要知识点分析 |
| 5 月 31 日 | 10:00-12:00 | Python 语言概述 Python 的基本语法 |
| | 14:00-17:00 | Python 的数据结构 Python 的控制结构 Python 的函数 Python 的迭代器和生成器 |
| 6 月 1 日 | 9:00-12:00 | Python 的类和对象 Python 程序的异常处理 Python 的包 Python 的内置标准库 |
| | 14:00-17:00 | 某音乐平台评论信息抓取 1. 获取歌曲 id 2. 抓取歌曲页面信息 3. 提取歌曲热门评论信息 4. 在 MySQL 数据库中构建歌曲评论信息数据表 5. 向数据库中导入歌曲评论数据 |
| 6 月 2 日 | 9:00-12:00 | web 服务器的 log 日志清洗 1. 提取访问 IP 地址 2. 提取访问的时间信息 3. 提取访问状态码、请求体的大小 |
| | 14:00-17:00 | 4. 数据清洗，封装提取的信息 |

| | | |
|------|------------|---------|
| | | 5. 分析统计 |
| 6月3日 | 9:00-12:00 | 考试及返程 |

四、授课师资（拟邀）

◇ 路老师

对外经济贸易大学副教授。中国科学院自动化研究所博士，新加坡国立大学博士后。主要研究方向为 Python 数据挖掘、用户建模等。在国内核心期刊和国际期刊和会议上发表论文 10 余篇，在数据挖掘和分析领域有着丰富的经验。

五、培训特色

权威的课程体系

- 以培养高等技术应用性专业人才为根本任务，以适应社会需求为目标、以培养技术应用能力为主线设计符合高职院校学生能力要求的培训内容



师资团队雄厚

- 院士、长江学者、领军企业知名专家云集
- 200+专业师资，来自北大、浙大、微软、腾讯等全国百余所高校、企业

配套实训平台

- 全程国内领先的理实结合实训教育云平台实践演练

授课形式

- 理论与实践结合，以数据为基础，实际问题为导向的案例实践性教学
- 精品小班授课，面对面指导

六、培训效果

借助博雅优秀的师资队伍、科学的课程设置、综合的实训平台，着重讲授 Python 语言，通过案例教学，帮助学员学习 Python 语言及前沿的网络爬虫、数据清洗技术，通过本次培训学员掌握以下技能：

- 掌握编程语言 Python 的基本语法、环境搭建、包的使用等内容；以真实数据的采集、清洗为案例和主线，掌握大数据处理的关键技术和流程；
- 具备大数据专业核心课程大数据的 Python 工具、数据采集与清洗课程的教学能力和案例设计能力，提升其科研能力；
- 了解高校大数据人工智能专业的实训内容，本次为每位参训学员提供大数据教学实训平台试用账号、课程的相关资源，丰富已开设大数据专业院校课程体系；
- 帮助参加培训的学员快速建立对大数据课程的整体性认识，了解大数据专业课程体系、人才培养方案，成为开发大数据专业课程及设置人才培养方案的优秀教师；

通过本次培训，为学校打造大数据领域优秀师资队伍，加快其大数据专业课程体系构建、师资队伍培养、配套实训平台建设及科学研究发展，促进江苏省大

数据教学水平不断提升。

七、报名须知

报名截止时间：2019年5月26日

报名流程：

- 1、咨询：报名咨询
- 2、报名信息填写：扫描右侧二维码填写报名信息
- 3、通知：工作人员核对信息，确认准确无误后向学员发送报名成功通知
- 4、缴费：学员现场刷卡或将学费汇至主办方账号，并把汇款底单邮件发送给王老师（mingyuew@cookdata.cn）



汇款信息如下：

收款人名称：北京博雅大智慧科技有限公司

开户银行：招商银行股份有限公司北京首体支行

银行账号：110929635110801

汇款形式（人民币）：请务必在备注栏里注明“江苏师资研修+院校/单位名称+姓名”

发票提示：发票类型为“培训费”，由北京博雅大智慧科技有限公司开具

- 5、开课通知：开课前1周工作人员将开课通知发送至学员邮箱，注意查收
- 6、报到：学员按“开课通知”的时间、地点准时报到，报到时携带1张2寸照片

八、培训认证

培训完成后用平台进行测试，考核合格，颁发北京大数据研究院博雅大数据学院大数据讲师/大数据分析师初级证书。



九、其他

➤ 联系方式

报名咨询： 王老师 15210681638 mingyuew@cookdata.cn
 王老师 18601990086 jingwciw@cookdata.cn

➤ 注意事项

各位学员请自带笔记本电脑

北京大数据研究院博雅大数据学院
(北京博雅大智慧科技有限公司)

2019年5月5日

附件一：企业介绍



北京大数据研究院于 2015 年 8 月 27 日，在北京市委市政府的支持与指导下，由中关村管委会、海淀区政府、北京大学、北京工业大学四方共同筹建。北京大数据研究院是国内首个整合了政府、大学和市场三方面资源的大数据研究机构，目标是吸引国际一流大数据研究人员来京发展，用五到十年的时间，建成国际一流的大数据教育、科研创新和产业化平台，成为中国乃至世界大数据产业发展的一面旗帜。



博雅大数据学院是由北京大数据研究院成立的专注于大数据教学产品研发、服务及教育培训的机构。两年来建立了完整、专业的本科、高职大数据课程体系、综合实训体系和政企定制化培训业务。通过聚焦数据科学、大数据和人工智能等前沿技术，培养大数据技术、应用及管理人才。目前已服务 80 多家企业、120 多所院校、超过 25000 多名全国院校师生及大数据从业者。

博雅拥有超过百名专业研发人员以及众多提供大数据优质授课服务的专家和教师。目前已经在国内大数据教育及培训行业中处于较为领先的地位，得到了业界的高度认可

往期培训



2016年-“大数据分析的模型与算法”暑期班



2017年-“大数据分析的理论与应用”暑期班



2018年-“数据科学与人工智能”暑期班



2018年-“数据科学”暑期特训

专项师资培训项目



2016年-第1期“数据科学导引”



2017年-第2期“数据科学与大数据技术”



2017年-“数据科学与大数据技术”



2017年-内蒙古医科大学数据科学专项培训



2017年-“大数据分析的Python基础”



2017年-“数据采集与网络爬虫”



2017年-“数据清洗技术与工具”



2017年-“大数据分析的原理与技术”



2018年-“大数据技术与应用”



2018年-湖北工业大学专项师资培训

大数据+教育



2017年-国家信息中心
大数据与干部教育



2017年-中国电信
大数据应用专项



2017年-中国电信云计算
分公司专场

大数据+通信



2017年- 中国电信
Spark实战专场



2018年-中国电信高管
人工智能专场



2018年-中国电信
星火计划专场

大数据+MBA



2017年-2018年长江学院
科技X商业MBA课程

大数据+智能制造



2017年-京东方集团
大数据技术与应用专场

大数据+IT



2017年-文思海辉
大数据分析的原理与技术

大数据+金融



2017年-金盟大数据管理课程
前沿技术与应用



2017年-招商银行
特征工程专场



2018年-Python数据分析
及机器学习应用

附件二：大数据技术与应用专业师资培训课程

课程体系



课程详情

专业基础课

1 大数据技术与应用导论

本课程系统介绍大数据的基本概念和发展历史，大数据的处理流程及核心技术，并通过实操案例帮助大家理解大数据在金融、教育、健康医疗等领域的应用。

课程大纲

| | |
|--------------|---------------|
| 1.大数据技术与应用概论 | 5.数据可视化 |
| 2.大数据的采集与清洗 | 6.大数据行业应用案例I |
| 3.大数据的存储与管理 | 7.大数据行业应用案例II |
| 4.大数据分析 | 8.大数据面临的挑战与未来 |

建议课程周期：2天

经典实操案例：

- 使用K近邻模型模型构造糖尿病诊断模型
- 客户分群分析
- 学生成绩预测

课程培养定位：大数据研发工程师方向、大数据分析师方向

2 大数据的Python基础

本课程系统介绍Python语言的基本内容和特性，帮助学员掌握Python语言基础，学会使用Python进行简单的数据分析。

课程大纲

| | | |
|---------------|------------------|-----------------|
| 1.Python语言概述 | 5.Python的函数 | 9.Python的包 |
| 2.Python的基本语法 | 6.Python的迭代器和生成器 | 10.Python的内置标准库 |
| 3.Python的数据结构 | 7.Python的类和对象 | 11.Python应用案例 |
| 4.Python的控制结构 | 8.Python程序的异常处理 | |

建议课程周期：3天

经典实操案例：

- 使用Python生成随机数序列
- 用类的方法构造矩阵
- 从迭代对象中选取不确定个元素
- 用Python实现对象运算符的重载
- Python文本分析实践

课程培养定位：大数据研发工程师方向、大数据分析师方向

3 大数据的Java基础

本课程主要介绍Java编程语言的基本内容，理解Java语言的特性，使学员具备编写Java应用程序的能力。

课程大纲

| | | |
|------------|---------|---------|
| 1.Java语言概述 | 4.数组 | 7.常用核心类 |
| 2.基本语法 | 5.类与对象 | 8.输入与输出 |
| 3.流程控制语句 | 6.继承与多态 | 9.异常处理 |

建议课程周期：5天

经典实操案例：

- Java连接数据库 (JDBC)
- 线性回归
- Java网络爬虫
- 利用Java对数据进行基本信息统计

课程培养定位：大数据研发工程师方向、大数据分析师方向

课程详情

专业核心课

4 Python数据分析

本课程系统介绍利用Python进行数据分析的基本内容和方法，通过对实际数据进行分

课程大纲

| | |
|--------------|-------------|
| 1. 数据分析概述 | 5. Pandas进阶 |
| 2. NumPy基础 | 6. Pandas绘图 |
| 3. NumPy通用函数 | 7. 实战案例 |
| 4. Pandas基础 | |

建议课程周期：3天

- 经典实操案例：
- 利用Pandas进行贷款人群分析
 - 基于Pandas绘图机制的医疗保险费用分析
 - 基于Numpy的联通股票分析
 - 马纹痛数据可视化

课程培养定位：大数据研发工程师方向、大数据分析师方向

5 数据可视化

本课程主要介绍数据可视化的理论和工具，以便了解各种图表的作用，熟练使用

课程大纲

| | |
|----------------|-----------------|
| 1. 数据可视化基础 | 4. Matplotlib画图 |
| 2. Excel可视化 | 5. Plotnine画图 |
| 3. Python数据预处理 | 6. 数据可视化案例 |

建议课程周期：2天

- 经典实操案例：
- Plotnine画图：Tips数据集的Plotnine可视化分析
 - 动画喜剧片的IMDB评分影响因素分析可视化
 - Matplotlib画图：学生成绩及影响因素的可视化分析
 - 共享单车使用量影响因素分析可视化

课程培养定位：大数据研发工程师方向、大数据分析师方向

6 Hadoop基础

本课程着重介绍大数据领域常用的计算框架Hadoop，内容包括：Hadoop搭建、文件系

课程大纲

| | |
|---------------------|-------------------|
| 1. Hadoop概述以及集群搭建 | 4. 分布式协调ZooKeeper |
| 2. 分布式文件系统HDFS | 5. 分布式查询引擎Hive |
| 3. 分布式计算框架MapReduce | 6. 分布式数据库Hbase |

建议课程周期：3天

- 经典实操案例：
- 1. Mahout银行客户聚类
 - 2. 室内网络评估系统优化
 - 3. 垃圾短信过滤效果提升
 - 4. Hbase性能调优实战
 - 5. 信用卡反欺诈实战

课程培养定位：大数据研发工程师方向

7 大数据行业应用导论

本课程从业务和模型两个角度解读大数据技术如何助力各行业发展，并解决其中的应用问题。

课程大纲

| | |
|----------|------------|
| 1. 金融大数据 | 3. 健康医疗大数据 |
| 2. 交通大数据 | 4. 电信大数据 |

建议课程周期：2天

课程培养定位：大数据研发工程师方向、大数据分析师方向

课程详情

专业选修课

8 Spark分布式平台

本课程系统介绍Spark的原理、算法，学会使用Spark操作数据库和实现机器学习模型。

课程大纲

| | | |
|--------------|---------------|--------------------|
| 1. Spark简介 | 3. Spark Core | 5. Spark Streaming |
| 2. Spark集群部署 | 4. Spark SQL | 6. Spark Mllib |

建议课程周期：2天

- 经典实操案例：
- 利用Spark Mllib预测房价
 - Spark Shuffle调优实战
 - 利用Spark Mllib实现协同过滤推荐

课程培养定位：大数据研发工程师方向

9 数据采集与清洗

本课程主要包括两个部分，一个部分为数据采集，第二个部分为数据清洗。通过学习掌握Python网络爬虫工具的使用方法，了解数据清洗的具体策略和方法。

课程大纲

| | | | |
|--------------|---------------|----------|-------------|
| 1. 初识网络爬虫 | 3. Scrapy爬虫框架 | 5. 数据预处理 | 7. 基于数据库的清洗 |
| 2. 网络基础与数据解析 | 4. 数据清洗基础 | 6. 数据转换 | 8. 数据清洗工具实战 |

建议课程周期：4天

- 经典实操案例：
- BasketBall数据采集
 - 爬取豆瓣电影评论
 - 爬取InfoQ博客技术文章

课程培养定位：大数据分析师方向

10 大数据的数学基础

本课程着重介绍了大数据的数学基础：微积分、概率论与数理统计以及线性代数的相关知识。帮助学员掌握基本的理论，提高学生的数学素养，培养学生的数学思维，从而有助于学生

课程大纲

| | | | |
|----------|--------------|-------------|---------------|
| 1. 函数 | 4. 行列式与线性方程组 | 7. 线性规划 | 10. 参数估计和假设检验 |
| 2. 极限与连续 | 5. 矩阵 | 8. 概率和随机变量 | 11. 数理统计基础 |
| 3. 微积分 | 6. 向量空间和特征向量 | 9. 数理统计基础知识 | |

建议课程周期：3天

- 经典实操案例：
- 利用奇异值分解进行图像压缩
 - 利用最小二乘法拟合一元直线和曲线
 - 非负矩阵分解实现及应用

课程培养定位：大数据分析师方向