



第5届全国高校大数据与人工智能教学研讨会

2022.05.13-2022.05.14 中国·厦门



大会官网

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位： 厦门大学 XIAMEN UNIVERSITY  天津大学 Tianjin University  中南大学 CENTRAL SOUTH UNIVERSITY  西南财经大学 SOUTHWEST UNIVERSITY OF FINANCE AND ECONOMICS  福州大学 FUZHOU UNIVERSITY  华侨大学 HUAQIAO UNIVERSITY  集美大学 JIAMEI UNIVERSITY  福建技术师范学院  人民邮电出版社 POSTS & TELECOM PRESS

协办单位： 美林数据 MERITDATA  TRANSWARP 星环科技  海豚实验室  山东省计算中心 SHANDONG COMPUTER SCIENCE CENTER  荔峰科技 LIFENG  Bai 百度  商迪3D web3di.com



面向数据科学的Python语言教学及课程思政建设

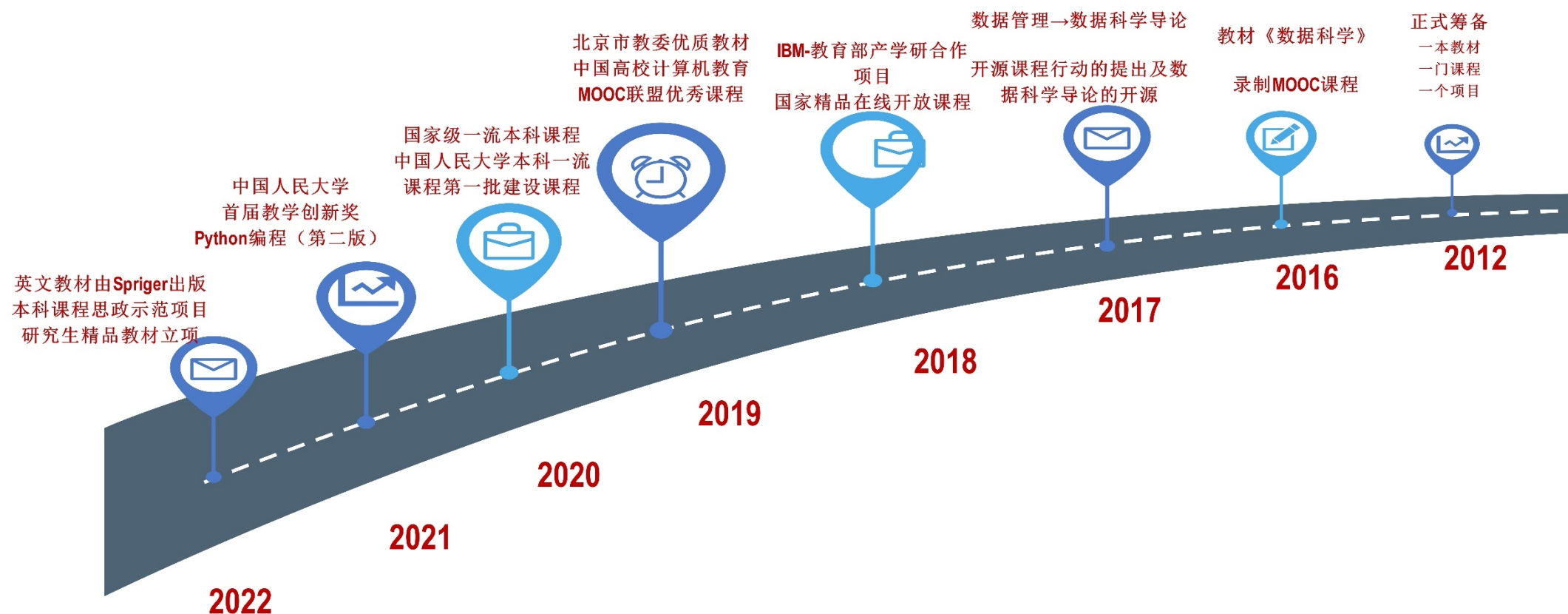
朝乐门

中国人民大学





《数据科学导论》课程的建设历程





提纲



Py4DS教学
中存在的问题



几点体会



课程思政建设





1. Python教学中存在的问题

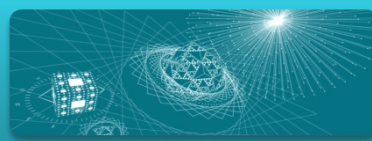




Python教学的误区或困境



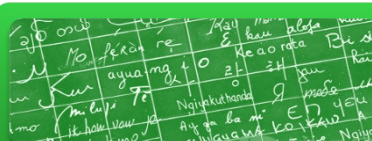
“将Python当作Java/C来教（或学）”，换一个“新语言”来讨论“老问题”



数类专业与计类专业中选用的Python教材没什么区别



只关注语法，不关注背后的思想和逻辑



“多数学生都已有C/Java等语言的基础，Python属于“第二外语”



“先讲知识点，后练习代码”式老套路，主次颠倒；



Pythonic Coding

Java

```
• Public Static void main(String arg[])  
{  
  int sqr[] = new int[100];  
  for(int i=0; i<100; i++){  
    sqr[i] = (i+1)*(i+1);  
  }  
  for(int i=0; i<100; i++){  
    System.out.println(sqr[i]);  
  }  
}
```

Python

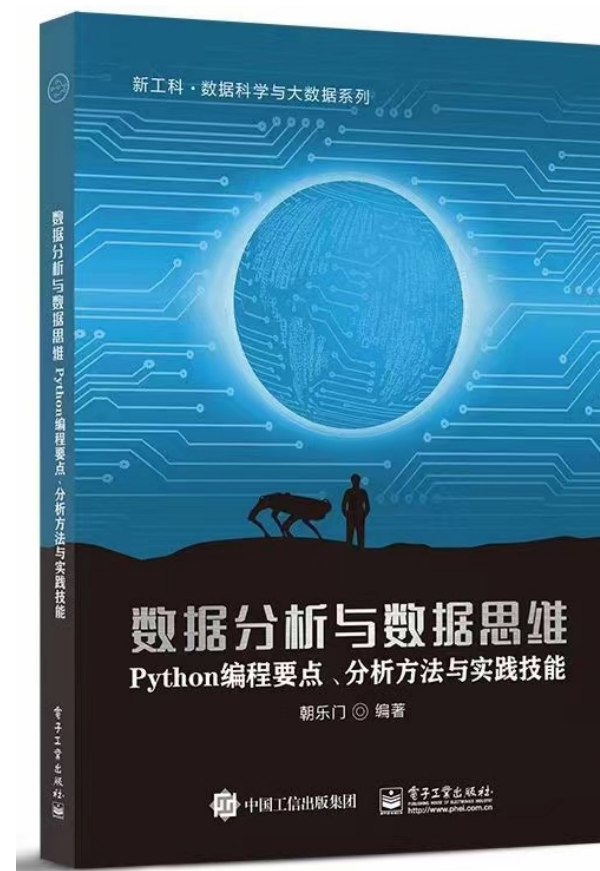
```
• sqr = [map(lambda a: a**2, range(1,101))]  
  print(sqr)
```



Pythonic Coding

```
for i in range(len(food)):  
    print(food[i])
```

```
for piece in food:  
    print(piece)
```





如何写出高质量的Python代码

写Python代码的两个境界

- 能run就ok?
- Pythonic coding

Python的Zen

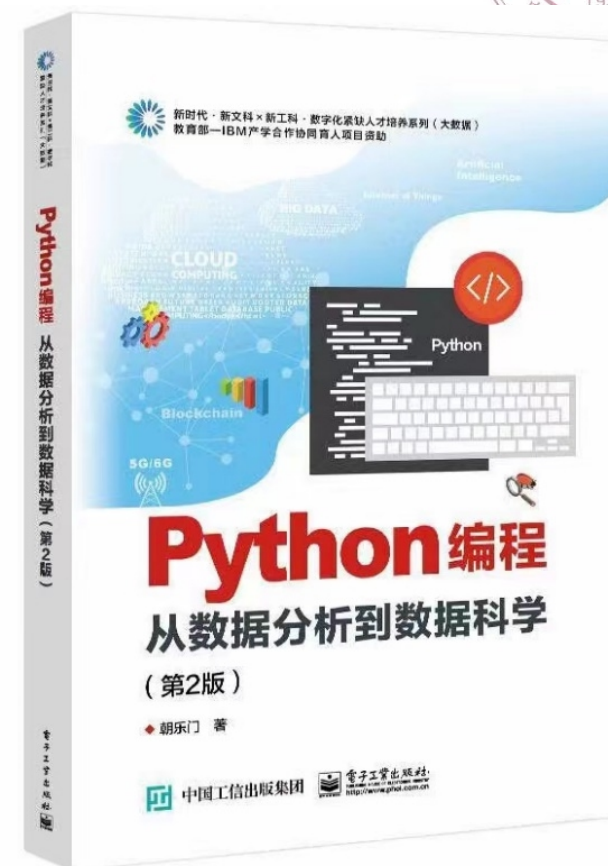
- import this

PEP8-Style Guide for Python Code

- adapted from *Guido's Python Style Guide*
- <https://legacy.python.org/dev/peps/pep-0008/>

Google Style Guides系列

- Google Python Style Guide
- <https://github.com/google/styleguide/blob/gh-pages/pyguide.md>





Python实践的几个重要问题

自己开发 or 第三方包？

编译式分析 or 交互式分析？

数据质量 or 数据模态？

标量化计算 or 向量化计算？

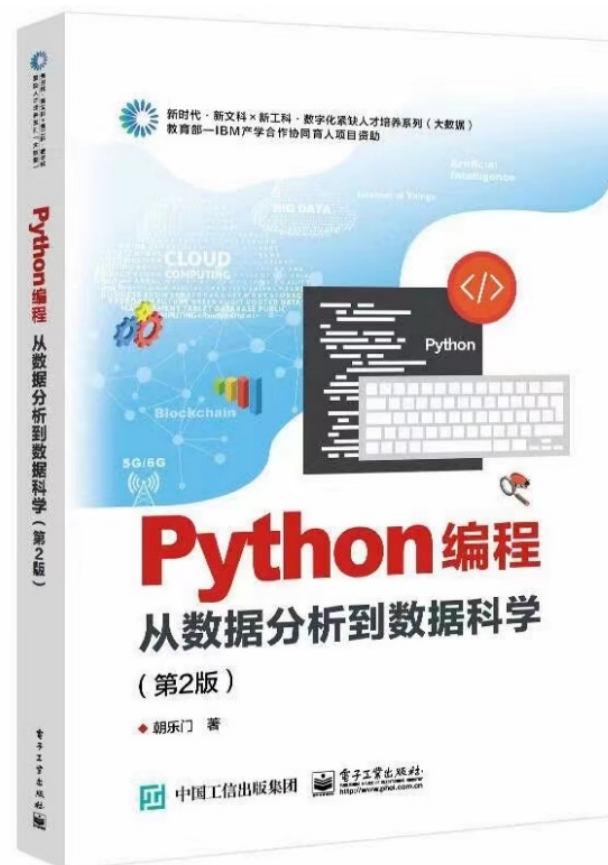
隐式索引 or 显式索引？

性能层次 or 功能层次？

验证性分析 or 探索性分析？

算法/模型的原理 or 算法/模型的应用？

数据工程 or 数据产品开发？





面向数据科学的Python语言教学内容

1. Python语法要点

- Python for DA
- 数据分析/数据科学项目中常用Python语法要点

2. 数据准备

- 随机数
- 多维数组
- 数据框
- 可视化
- Web爬取

3. 算法与模型

- 机器学习
- 统计学
- 图像处理
- 自然语言理解

4. 大数据分析

- Spark
 - MongoDB
 - MLib
- 如何用Python学习数据科学





2. 几点体会





什么是“金课”？

高阶性

- 知识、能力、素质的有机融合
- 培养学生解决复杂问题的综合能力和高级思维

创新性

- 课程内容要反映前沿性和时代性
- 教学形式呈现先进性和互动性
- 学习结果具有探究性和个性化

挑战度

- 老师备课和学生课下有较高要求
- 学生和老师一起，需要跳一跳才能够得着



(来源：教育部官网)



2.1 优质的课程内容是“王”

岗位面试题调研

- 全球DS领域岗位面试题调研
- 中文
- 英文
- 500/1930



招聘公告调研

- 全球DS领域招聘公告的调研
- 岗位职责
- 任职要求要求

课程调研

- 全球数据科学导论课程的建设现状
- 2012年，现状调研
- 2012-2020，跟踪研究

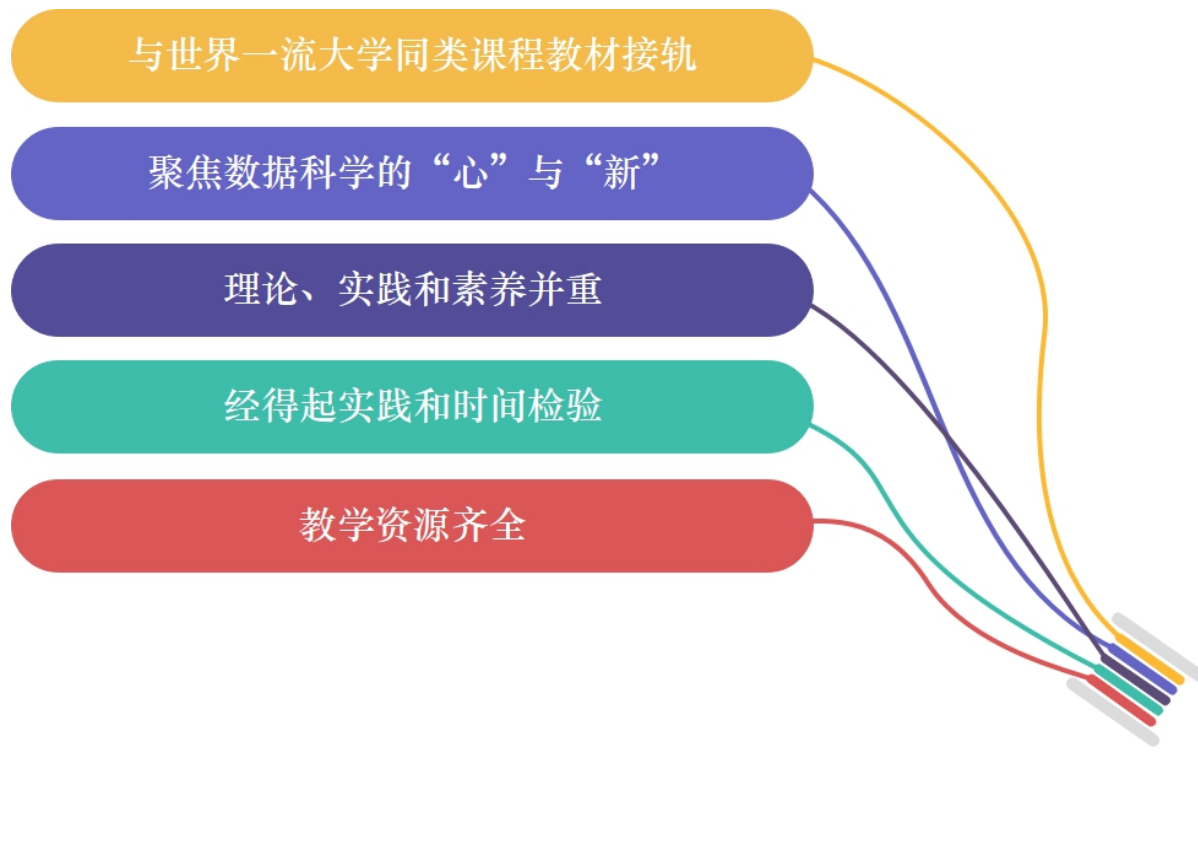


课程链调研

- DS专业特色课程的调研
- 未来数据科学家都在玩什么
- 新课程
- 特色课程

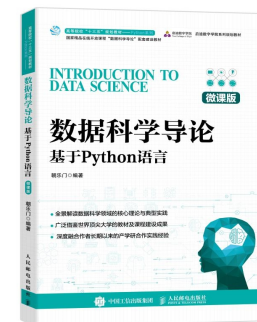


2.2 优质教材是主要抓手



数据科学
理论与实践（第2版）

清华大学出版社

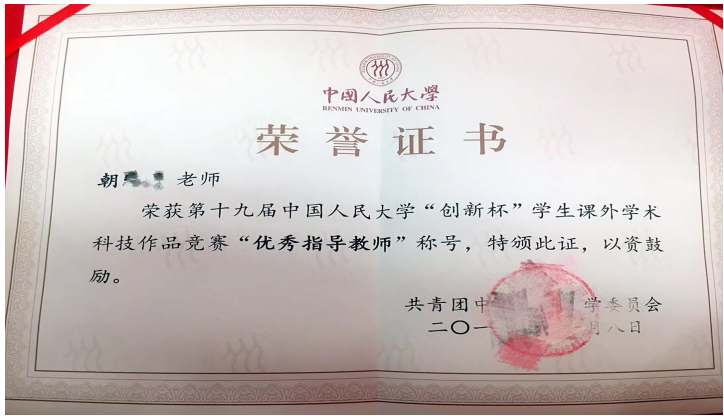




2.3 培育学生的学习兴趣和信心是关键



H奖 (15.35%)



“竞赛与开源项目是通往数据科学家的两条最佳捷径。”
——摘自《数据科学理论与实践》





2.4 教学内容从信息到知识/智慧的转变

**Where is the Life we
have lost in living?
Where is the wisdom we
have lost in knowledge?
Where is the knowledge
we have lost in
information?**



(来源: T.S.Eliot, The Rock , 1934)



2.5 配套平台与资源的建设

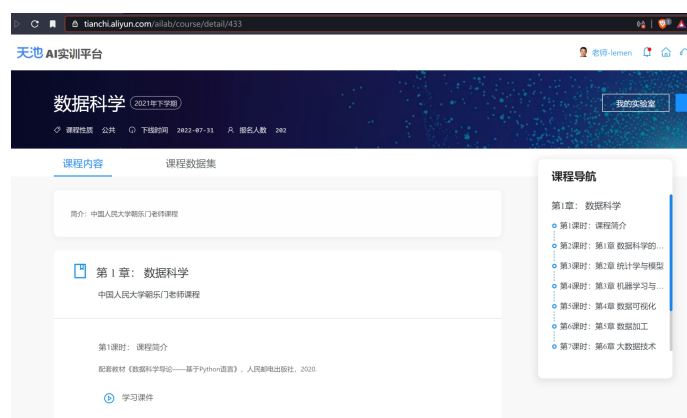
朝老师推荐的【数据科学】领域经典文章



朝老师推荐的【面向数据科学的Python编程】系列文章



阿里云 TIANCHI 天池





2.6 慕课的“灯下黑”及开源课程行动



教学大纲

新开课申请表及课程简介

讲义PPT

源代码（Python版）

原始数据

习题参考答案

教师参考书目

数据科学相关重要资源目录

支撑平台及作者联系方式

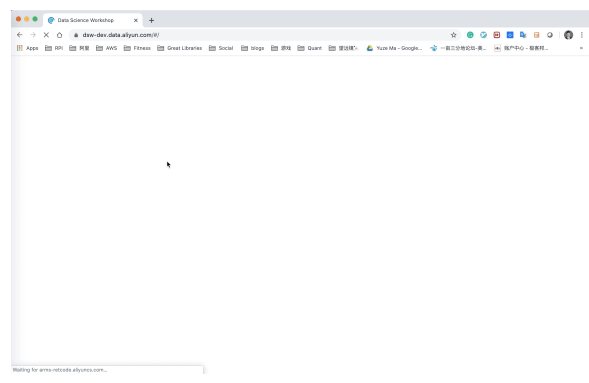
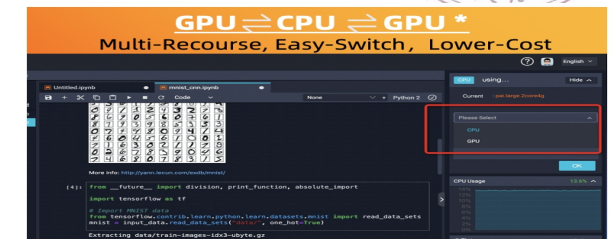
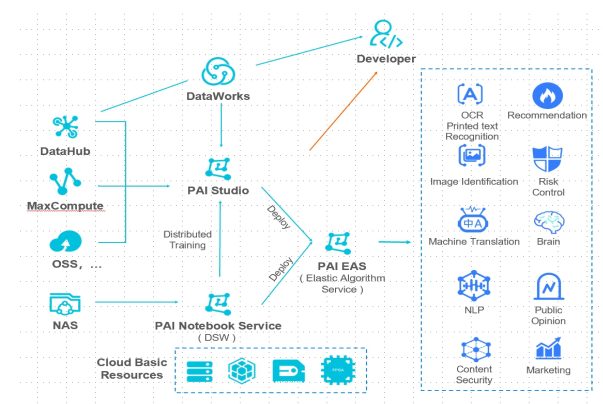
教材勘误信息表

GitHub地址





2.7 有效关联至其他优质在线资源



每门课的开始知识分享，最期待/最不希望这门课...?

2.8 以学生为中心的教学设计

指导/启发、研讨性教学



描述分析结果-简化格式

名称	样本量	平均值	标
第一组评分 组员: 胡开朗 (组长)、陈晓珑 演示内容: 线性回归	24	2.750	0.
第二组评分 组员: 严奕聪 (组长)、李若兰 演示内容: 逻辑回归	24	2.500	0.
第三组评分 组员: 邓宁远 (组长)、刘阳 演示内容: 泊松回归	24	2.042	0.
第四组评分 组员: 任一萌 (组长)、张梦依 演示内容: 方差分析	24	2.000	0.
第五组评分 组员: 胡砚勇 (组长)、李宜芳 演示内容: 时序分析	24	2.208	0.
第六组评分 组员: 许卉 (组长) 演示内容: PCA	24	2.292	0.

▼ 专题【数据科学的基础理论】 ▶ 10. 如何继续学习

如何继续学习专题知识



学好本章的重要意义

- 正确理解研究目的、理论体系与基本原则等核心问题是学好数据科学的第一步，防止学习的盲目性和效率低的重要前提，更是成长为数据科人才的必要条件。

继续学习方法

- 本章介绍了数据的科学管理基础理论，但数据的科学管理是一门新兴学科，相关理论也在动态变化中。因此，在掌握本章内容的基础上，学会自己跟踪数据科学领域的重要期刊、会议、图书、专家的方法尤为关键。

提醒及注意事项

- 学习数据科学中应注意四个基本问题——数据科学的四则运算原则，如下图所示。





2.9 积极探索教学改革与同行交流



朝乐门.数据科学导论的课程设计及教学改革[J].计算机科学,2020,47(07):1-7.

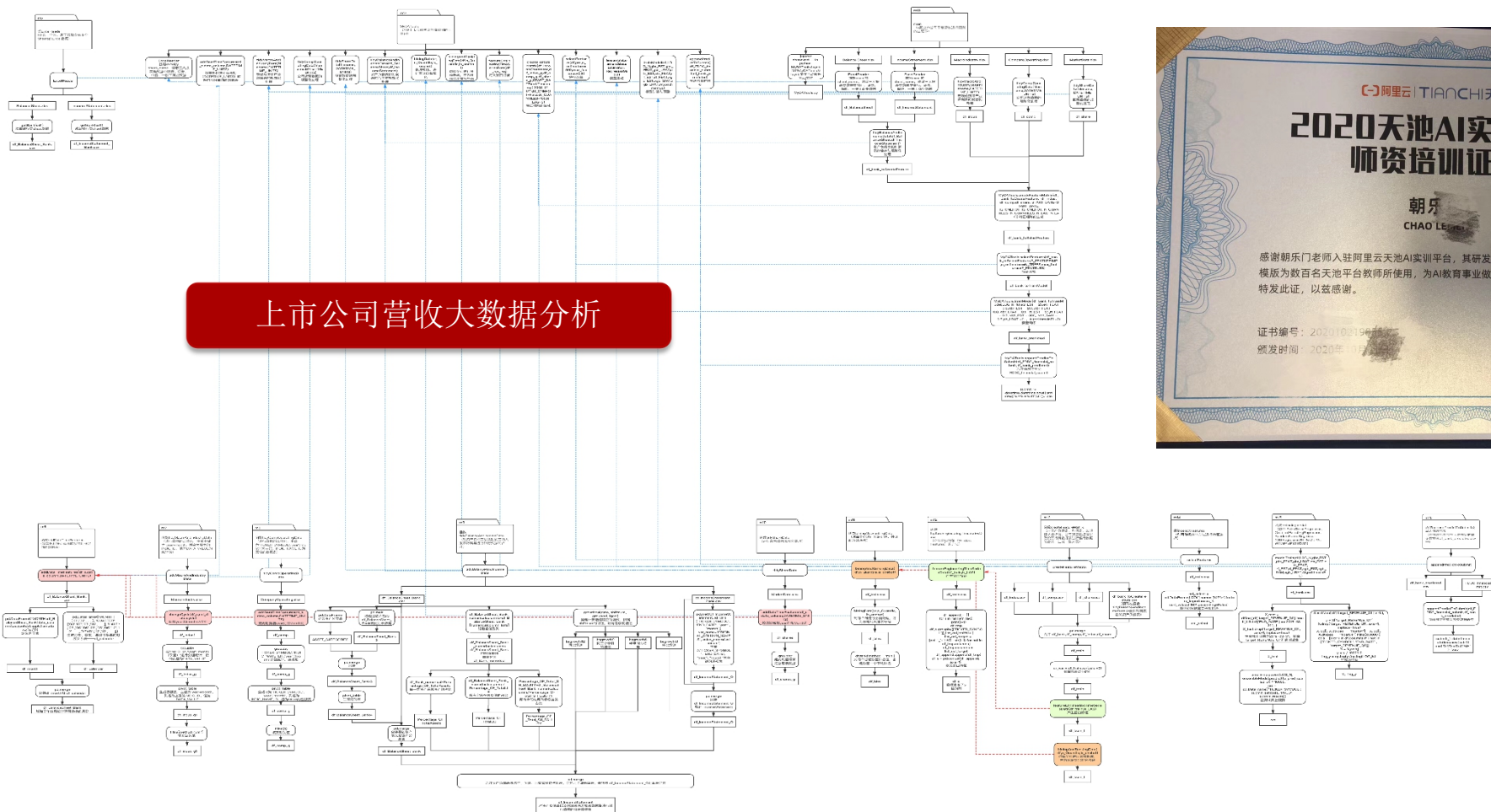


朝乐门等.全球数据科学课程建设现状的实证分析[J].数据分析与知识发现,2017,1(06):12-21.



朝乐门等.数据科学与大数据技术专业特色课程研究[J].计算机科学,2018,45(03):3-10.

2.10 集中精力研发课程综合实训案例





3. 课程思政建设



高等学校课程思政建设指导纲要



中华人民共和国教育部
Ministry of Education of the People's Republic of China

当前位置: 首页 > 公开

信息名称: 教育部关于印发《高等学校课程思政建设指导纲要》的通知
信息索引: 360A08-07-2020-0015-1 生成日期: 2020-06-01 发文机构: 中华人民共和国教育部
发文字号: 教高〔2020〕3号 信息类别: 高等教育
内容概述: 教育部印发《高等学校课程思政建设指导纲要》。

教育部关于印发《高等学校课程思政建设指导纲要》的通知

教高〔2020〕3号

各省、自治区、直辖市教育厅（教委），新疆生产建设兵团教育局，有关部门（单位）教育司（局），部属各高等学校、部省合建各高等学校：

《高等学校课程思政建设指导纲要》已经教育部党组会议审议通过，现印发给你们，请结合实际认真贯彻执行。

教育部

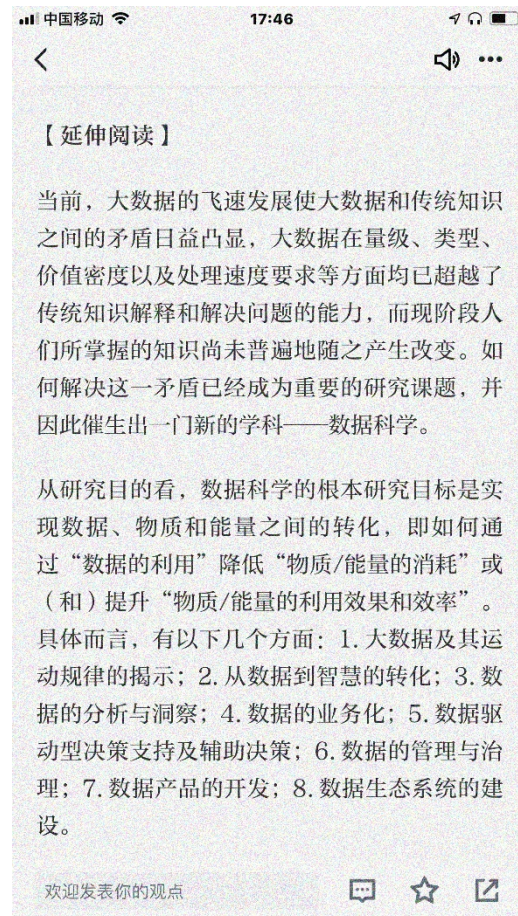
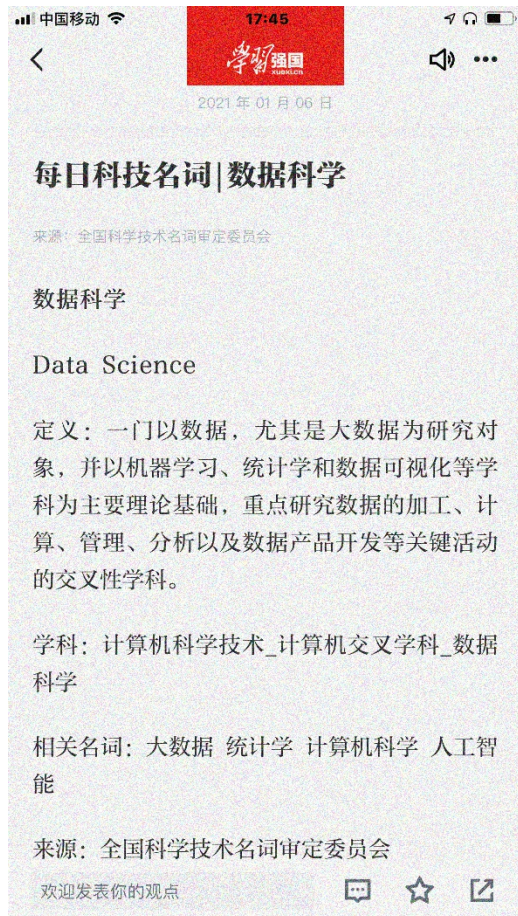
2020年5月28日

来源：
教育部官网



课程思政融入数据科学课程教学

数据科学是一门以数据，尤其是大数据为研究对象，并以机器学习、统计学和数据可视化等学科为主要理论基础，重点研究数据的加工、计算、管理、分析以及数据产品开发等关键活动的交叉性学科。





课程思政的几点体会

<p>增强“四个意识”、坚定“四个自信”、做到“两个维护”</p>	<p>思政课程与课程思政同向同行，产生协同效应</p>	<p>加快中国特色、世界一流的教材建设</p>	<p>研发基于经典算法的Python数据科学典型案例</p>	<p>数据集与研究问题的选择很关键，重视中国问题，中国解决方案</p>



欢迎您加入数据科学虚拟教研室

联系方式:

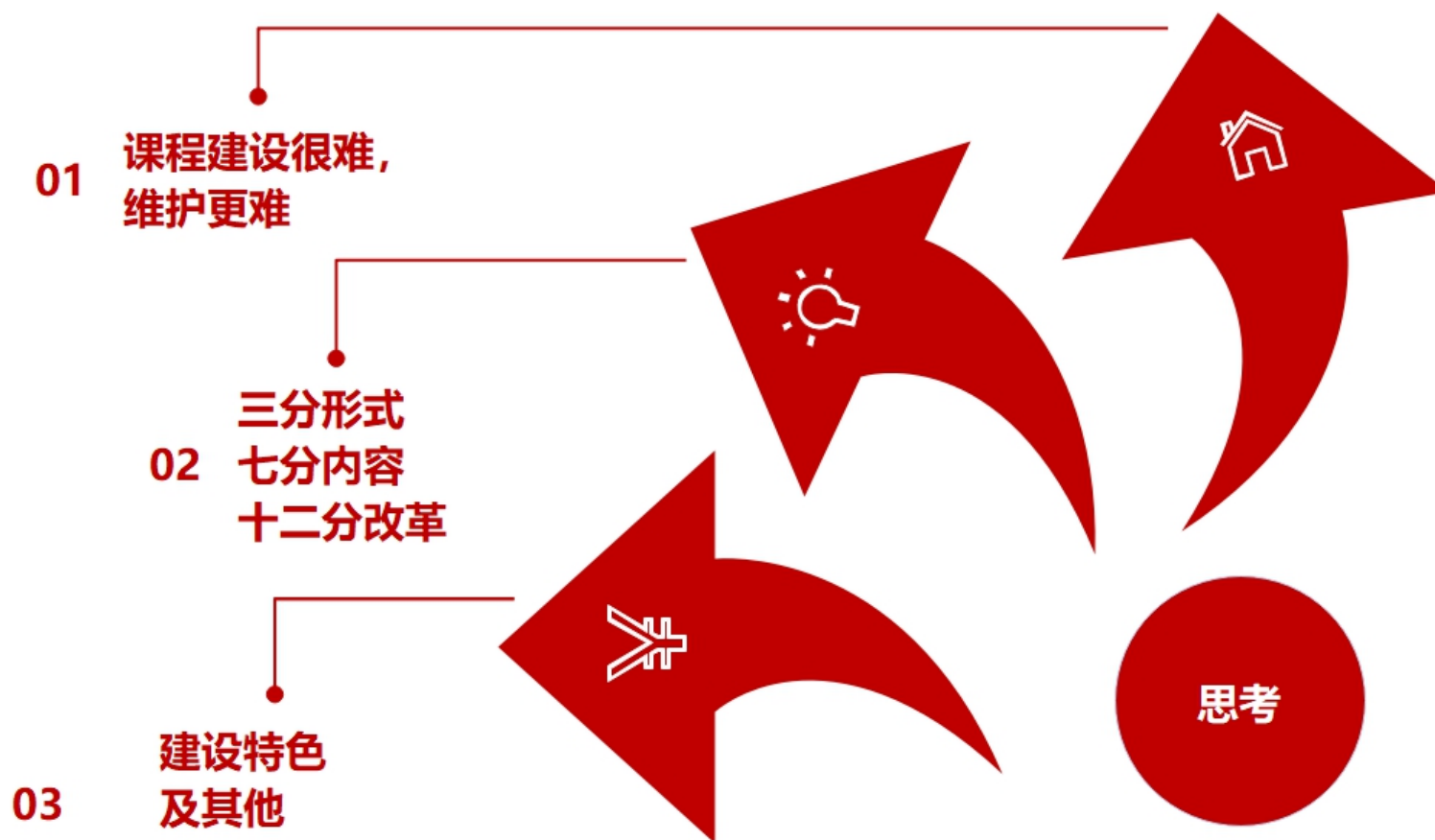
chaolemen@ruc.edu.cn



【数据科学虚拟教研室】公众号



结束语





【数据科学虚拟教研室】
公众号



主讲人联系方式



主讲人微信