

# 数据科学工具

[√]科学、技术与环境

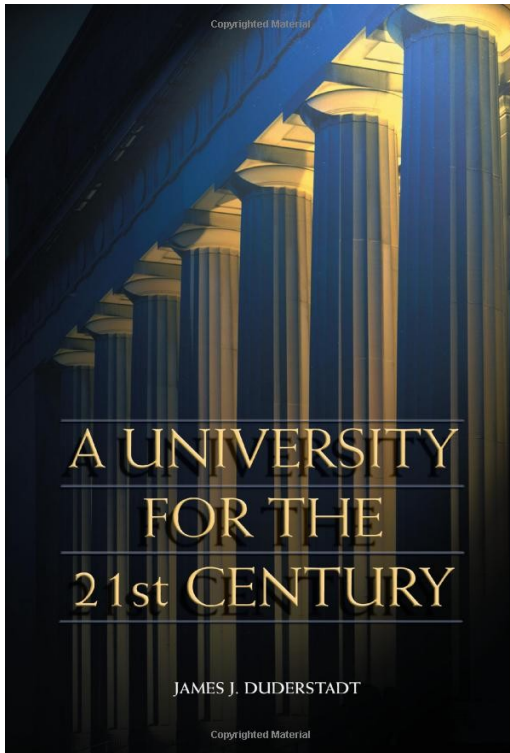
李丰

统计与数学学院

[feng.li@cufe.edu.cn](mailto:feng.li@cufe.edu.cn)

<http://feng.li/>

# 21世纪的大学需要什么样的通识教育？



“……尽管难以给通识教育下定义，它的实现也极具挑战性，但这难以捉摸的通识教育的目标可能依然是使学生为终身学习和**变化的世界**做好准备的最好途径。”——现代教育学家密歇根大学校长Duderstadt

**数据以及其所涉及的科学正在改变这个世界！**

**我们的学生准备好了吗？**

# 一个真实的故事

【思享·中财·第128期】夏凡盛：中央财经大学 金融学院金融学学硕  
备考经验分享

原创：夏凡盛 CUFES享 5月29日

## 基本信息

姓名：夏凡盛

本科专业：文化与传媒学院广告学

本科成绩：85.77/100

报考院校专业：中央财经大学金融学学硕

初试成绩：总分388 = 政治65 + 英语一74 + 数学三107 + 801经济学142，初试排名第13

报考专业复试情况：分数线371，34人进入复试，最终录取27人，本人复试后综排第4



“觉得数据和统计真的非常有用和觉得数据和统计真的非常有用和神奇，特别是大三下参加了统数学院李丰老师的一个Python课程后，大部分空闲时间都花在这上面”

“大三下旁听了李丰老师的Python课在复试面试中帮了大忙”

“自我介绍还提到了一下自己对统计有兴趣参加过一个学校的Python课，老师就问了因子分析具体如何操作和Python常用的数据分析库有哪些以及因子分析在Python要怎么实现，因为自己对这些都比较熟，所以老师对回答也都挺满意的，学的东西总算派上用场了。”

# 课程简介

- 本课程为面向我校一年级（下）及以上对数据科学感兴趣的本科生开设的一门现代数据科学相关工具的通识核心课程。
- 本课程设计模块独立，不要求学生有编程基础。

# 数据科学工具课程教学目标

从数据到价值的  
实践能力

依靠数据的科学  
决策能力

案例  
教学

熟悉数据科学分析平台

理解数据科学中的数据

应用数据科学工具

# 教学内容

章节	教学形式	章节	教学形式
第一章：数据科学核心价值	课堂教学与案例式教学	第五章：数据可视化	体验式教学
第二章：现代数据生产环境与工具	课堂教学与互动式编程	第六章：海量数据分布式计算基础	课堂教学与互动式编程
第三章：创建自己的大数据仓库	课堂教学与案例式教学	第七章：大数据价值实现	课堂教学与互动式编程
第四章：非结构化数据处理基础	课堂教学与案例式教学	第八章：数据科学应用案例	案例式教学

# 数据科学工具教学环境

## 虚拟实验环境



### 数据科学平台

学生无需在自己电脑部署这些平台，便于案例能够迅速进行



### 虚拟云桌面

通过客户端或者网页实时查看实验流程，系统实时反馈实验进度



### 前期开发案例库

学生通过完整的案例学习，优秀案例会补充到已有案例库

# 数据科学工具课程前期案例库建设

方向	案例名称
财经文本数据	<ul style="list-style-type: none"><li>✓ 基于法律文书的上市公司纠纷研究</li><li>✓ 利用大数据分布式计算从海量文本数据中发掘新闻情绪与股票收益的关联</li></ul>
数据科学与算法	<ul style="list-style-type: none"><li>✓ 亚马逊在线书店新书预测策略</li><li>✓ 基于物品的协同过滤算法的推荐系统</li><li>✓ 基于气候信息的零售商店药品销售与推荐</li></ul>
信用风险度量与管理	<ul style="list-style-type: none"><li>✓ P2P平台信用风险评估</li><li>✓ 基于文本与财经数据的国有企业债务违约测度</li></ul>
数据科学聚类	<ul style="list-style-type: none"><li>✓ 基于DSBSCAN对科技股票涨跌情况聚类</li><li>✓ 基于MapReduce的股票聚类分析</li><li>✓ 基于欧洲五大联赛2014-2015赛季数据的线性判别分析</li><li>✓ 基于KNN模型的游戏舰队分类</li><li>✓ 基于MapReduce对某大学女生支出行为的聚类分析</li></ul>
.....	<ul style="list-style-type: none"><li>✓ 我国零售业上市公司财务绩效研究——基于MapReduce的并行主成分分析</li><li>✓ 应用商店APP应用流行趋势</li><li>✓ 基于树的方法对白酒质量分类并行计算实现</li></ul>



# 科学素养与一体化的知识结构

- 数据科学不断改变业界、学者、学生的思维模式。加强学生数据科学工具使用的培养，既是对大数据时代的融入，也是未来的方向。
- 本通识课程强调以实际应用为导向的**知识思维和学习模式**，避免初学者学习目的盲目的问题，又有效地平衡和数据科学内容繁复与学习精力有限的问题。
- 本通识课程在方法论上体现利用数据科学知识与专业知识结合，培养学生**跨界融合的创新精神**。
- 本通识课程在科学实践中，引入现代交互式教学模式，学生可以不断校正学习模式，提高发现、**利用一手数据探究问题本质的能力**。

# 来自学生的声音

- 2017年3月底，联合教学技术服务中心为全校师生举行为期6周共计18学时关于数据科学工具的入门培训。
- 参加现场培训和线上直播的师生总计881人次，其中教师83人次。
- 摘录部分教师和学生的反馈如下：

“感觉在学校能有这样的培训真的很惊喜，一开始自己对数据科学只是比较好奇，现在越来越感兴趣了，即使课程结束应该自己也会继续学习相关方面的内容”

“课程的选题非常热门，很有吸引力。这种课程的形式适合入门，今后有更深一层需要时，可以再自学。多媒体形式丰富、很炫酷。”

“让我对现代数据的处理方法有了新的理解，为我提供了新的学习方向”

“觉得世界好大，非常感谢老师！”

“希望老师能够把这门课开成任选或者是限选课”

“对Python和大数据有了初步的了解，希望以后可以多多开设这种课程。”

授之以渔而胜于鱼

