

中央财经大学通识核心课程建设项目教学大纲

课程名称（中）	数据科学工具				学分	学时
课程名称（英）	Tools for Data Science				2	36
开课单位	统计与数学学院	课程负责人	李丰	职称	讲师	
拟所属模块	<input type="checkbox"/> 语言、文学与艺术 <input type="checkbox"/> 历史、政治与社会 <input type="checkbox"/> 哲学、心理与生活 <input checked="" type="checkbox"/> 科学、技术与环境 <input type="checkbox"/> 生命、地球与宇宙					
考核方式 (划“√”)	开卷（ ） 闭卷（ ） 论文（√） 口试（ ） 其他（ ） 可多选					
	过程考核成绩： 60%		期末考核成绩： 40%			
课程介绍【授课对象、开设目的、教学内容（具体到章）、教学方法、考核方式、教学特色等；字数 1000 字以内】						
<p>数据科学工具课程的开设是基于课程团队成员对本科生课程开设的一门全新的以数据科学为依托、利用计算机工具链获取一手数据资源、处理海量非结构化数据的通识类课程。这是一门定位于面向财经类学生的数据科学工具性课程。这与国际上数据科学相关课程主要面向计算机相关专业不同，降低了财经相关专业学生学习数据科学的门槛。其次，该课程可以作为其他财经专业课程的拓展课程，学生通过课程的学习，能够了解并掌握通过互联网大数据资源获取和分析一手数据的能力。再次，本课程为学生提供了分析海量非结构化数据的数据科学工具链，培养学生数据科学分析的基本素养。</p>						
教学目标【主要表述学生课程学习后在知识、方法、能力、内在素养等方面达到的水平；字数 300 字以内】						
<p>该课程的主要内容涵盖数据科学分析的基本工具并体现现代数据分析的基本思路。本课程首先从企业大数据实例出发，使学生更好的理解大数据环境下数据处理的特点以及学习大数据技术的必要性。然后该课程介绍实际应用中的大数据平台基础，授课内容主要包括：Linux 系统中的命令行、Python 基础、基于 Python 的网络数据抓取、基于 Python 的自然语言处理、非结构化数据分析、海量数据并行计算、Hadoop 分布式存储与计算。课程最后以案例分析的形式总结数据科学工具平台在实际问题中的应用。</p>						

教学内容及安排【理论课程内容具体到节或知识块，用分号断开；标“△”表示重点，“★”表示难点。课程设计、实习列出内容、安排、步骤和要求等。】

章号	教学内容	时间分配（学时）					小计
		讲课	实验上机	课堂讨论	课外实践	其他环节	
01	数据科学核心价值	2					
02	现代数据生产环境与工具	4	2				
03	创建自己的大数据仓库	4	1	1			
04	非结构化数据处理基础	4	1	1			
05	数据可视化	2	1	1			
06	海量数据分布式计算基础	2	1	1			
07	大数据价值实现	2	1	1			
08	数据科学应用案例	2		2			
合 计		22	7	7			

实验、实践项目（可选）

序号	实验（实践）名称	实验（实践）内容与要求	学时
01		Python 常用数据分析与机器学习模块	1
02		复杂文本数据案例与建模分析	1
03		利用 Python 解析财务报告数据	1
04		复杂数据展示文本数据可视化	1
05		利用现代在线平台完成一个 MapReduce 实现	1
06		Hadoop Streaming 与 Python 数据建模	1
07		利用网络大数据分析新闻情感与股票波动	1

教材与参考书目			
书名	作者	出版社	出版年度
大数据分布式计算	李丰	中国人民大学出版社	2017
<i>Data Science from Scratch</i>	Joel Grus	O'Reilly Media	2015
<i>Web Scraping with Python</i>	Ryan Mitchell	O'Reilly Media	2015
课程考核			
考核方式	选择闭卷笔试、开卷笔试、口试、面试、答辩、课程论文、设计报告、实习报告、实验报告、操作技能考核、作品展示等。 课程论文		
成绩记载	期末考试、平时测验、课程作业、实验考核等成绩构成及所占比例 本课程是一门以应用和实践为主的课程，其考核方式将主要结合平时上机作业加期末考核为主。平时上机实践是学生顺利完成学习目标的必要手段，平时作业占50%。期末考核点为学生是否能够利用已学数据科学工具独立完成数据分析案例分析报告（自由选题）。合格的期末报告应体现学生对一手数据的独立获取能力和探索复杂数据价值的的能力。期末报告占40%，平时出勤及课堂表现占10%		