

## 《大数据技术及应用》本科课程教学大纲

## 一、课程基本信息

课程名称	(中文) 大数据技术及应用				
	(英文) Big Data Technology and Application				
课程代码	2055043	课程学分		3	
课程学时	48	理论学时	16	实践学时	32
开课学院	信息技术学院	适用专业与年级		计算机科学与技术专业 大学三四年级	
课程类别与性质	选修课	考核方式		考查	
选用教材	Python 数据分析与应用 (微课版), 丁菊玲著, 人民邮出版社, 2025 年 1 月			是否为马工程教材	否
先修课程	计算机导论				
课程简介	<p>大数据技术及应用是一门新兴的交叉性学科,是在信息技术领域和人工智能领域迅速兴起的计算机技术。数据挖掘技术面向应用,在很多重要的领域,数据挖掘都发挥着积极的作用。广大从事数据库应用与决策支持,以及数据分析等学科的科研工作者和工程技术人员迫切需要了解和掌握它。因此数据挖掘已经成为计算机专业及相关专业的重要课程之一。</p> <p>本课程为计算机专业学生的专业实践课程。本课程主要介绍数据挖掘的基本概念,原理、方法和技术。旨在通过一学期的学习,使学生理解数据挖掘的基本流程,掌握数据挖掘的基本理论和技术,熟悉数据挖掘成果的显示;掌握数据挖掘的基本方法,能熟练地应用数据挖掘技术对现实数据进行有效的分析;结合相关软件能从大量数据中获取有价值的信息。</p>				
选课建议与学习要求	本课程是适用于计算机类专业的专业实践课程,要求具有计算机导论的基础、基本编程能力和系统设计能力、对于数据库的基本操作能力和算法设计能力。				
大纲编写人	彭青松 (签名)		制/修订时间	2025 年 08 月	
专业负责人	戴志明 (签名)		审定时间	2025 年 09 月	
学院负责人	(签名)		批准时间		

## 二、课程目标与毕业要求

### (一) 课程目标

类型	序号	内容
知识目标	1	熟练掌握 Python 的基本语法与核心库（如 Pandas、NumPy、Matplotlib、Scikit-learn 等），能够高效处理金融数据的清理、转换、分析与可视化。
	2	掌握网络爬虫的基本原理与技术，包括静态网页采集（Requests、Beautiful Soup、lxml）与动态内容抓取（Puppeteer、Selenium），能够从多种数据源（如开放数据集、API、网页）中获取金融数据。
	3	使用 TA-Lib 计算常见技术指标（如 MA、RSI、MACD、布林带等），并基于这些指标设计交易策略（如趋势跟踪、均值回归等），掌握金融数据分析与策略优化的方法。
	4	了解金融数据的存储方式及其在大数据场景下的应用，掌握传统关系型数据库（如 MySQL）与分布式存储技术（如 HDFS、HBase、NoSQL 数据库）的使用方法，能够根据数据特性与性能需求选择合适的存储方案。
技能目标	5	数据采集与预处理能力，能够熟练运用网络爬虫技术从多种数据源中采集金融数据，并对数据进行清理、转换与预处理，确保数据质量。
	6	金融数据分析与策略设计能力，能够使用 TA-Lib 计算常见技术指标，并基于这些指标设计、优化与回测交易策略。同时，能够利用 Python 工具链（如 NumPy、Matplotlib、Scikit-learn）进行数据分析、可视化与建模。
	7	大数据存储与高效计算能力，能够根据数据特性与性能需求，选择合适的存储方案，并利用 Python 多进程技术实现高效的数据处理与并行计算，提升量化系统的性能。
素养目标 (含课程思政目标)	8	学生将深刻理解数据安全与隐私保护的重要性，遵守相关法律法规，在数据采集、存储与分析过程中恪守职业道德，杜绝非法爬取、滥用数据等行为。
	9	树立科技报国与创新驱动发展的使命感，学生将认识到金融科技对国家经济发展的重要性，激发科技报国的使命感，培养创新精神与实践能力，努力将所学知识应用于金融领域的自主创新，助力我国金融科技的国际化与竞争力提升。

### (二) 课程支撑的毕业要求

<p>L02 问题分析：能够应用数学、自然科学和工程科学的基本原理，识别、表达、并通过文献研究分析复杂工程问题，以获得有效结论。</p> <p>①具备对系统设计、软件开发等涉及到的复杂工程问题进行识别与判断，并结合专业知识进行有效分解的能力。</p> <p>④在充分理解专业知识的基础上，能够运用所学知识开展文献检索和资料查询。</p>
<p>L06 工程与社会：能够基于工程相关背景知识进行合理分析，评价专业工程实践和复杂工程问题解决方案对社会、健康、安全、法律以及文化的影响，并理解应承担的责任。</p> <p>②熟悉计算机专业领域相关的技术标准、知识产权、产业政策和法律法规。</p> <p>③能客观评价计算机应用项目的实施对社会、健康、安全、法律以及文化的影响。</p>
<p>L07 环境和可持续发展：能够理解和评价针对复杂工程问题的专业工程实践对环境、社会可持续发展的影响。</p> <p>①了解与本专业相关的职业和行业的生产、设计、研究与开发、环境保护和可持续发展等方面的方针、政策和法律、法规。</p> <p>②能正确认识并评价计算机科学在现实社会中应用的影响。</p>

### (三) 毕业要求与课程目标的关系

毕业要求	指标点	支撑度	课程目标	对指标点的贡献度
L02	①	L	掌握网络爬虫的基本原理与技术，包括静态网页采集 (Requests、Beautiful Soup、lxml) 与动态内容抓取 (Puppeteer、Selenium)，能够从多种数据源 (如开放数据集、API、网页) 中获取金融数据。	10%
			使用 TA-Lib 计算常见技术指标 (如 MA、RSI、MACD、布林带等)，并基于这些指标设计交易策略 (如趋势跟踪、均值回归等)，掌握金融数据分析与策略优化的方法。	30%
			了解金融数据的存储方式及其在大数据场景下的应用，掌握传统关系型数据库 (如 MySQL) 与分布式存储技术 (如 HDFS、HBase、NoSQL 数据库) 的使用方法，能够根据数据特性与性能需求选择合适的存储方案。	30%
			大数据存储与高效计算能力，能够根据数据特性与性能需求，选择合适的存储方案，并利用 Python 多进程技术实现高效的数据处理与并行计算，提升量化系统的性能。	30%
	④	M	熟练掌握 Python 的基本语法与核心库 (如 Pandas、NumPy、Matplotlib、Scikit-learn)	20%

			等), 能够高效处理金融数据的清理、转换、分析与可视化。	
			掌握网络爬虫的基本原理与技术, 包括静态网页采集 (Requests、Beautiful Soup、lxml) 与动态内容抓取 (Puppeteer、Selenium), 能够从多种数据源 (如开放数据集、API、网页) 中获取金融数据。	20%
			使用 TA-Lib 计算常见技术指标 (如 MA、RSI、MACD、布林带等), 并基于这些指标设计交易策略 (如趋势跟踪、均值回归等), 掌握金融数据分析与策略优化的方法。	30%
			了解金融数据的存储方式及其在大数据场景下的应用, 掌握传统关系型数据库 (如 MySQL) 与分布式存储技术 (如 HDFS、HBase、NoSQL 数据库) 的使用方法, 能够根据数据特性与性能需求选择合适的存储方案。	30%
L06	②	M	数据采集与预处理能力, 能够熟练运用网络爬虫技术从多种数据源中采集金融数据, 并对数据进行清理、转换与预处理, 确保数据质量。	50%
			学生将深刻理解数据安全与隐私保护的重要性, 遵守相关法律法规, 在数据采集、存储与分析过程中恪守职业道德, 杜绝非法爬取、滥用数据等行为。	50%
	③	M	金融数据分析与策略设计能力, 能够使用 TA-Lib 计算常见技术指标, 并基于这些指标设计、优化与回测交易策略。同时, 能够利用 Python 工具链 (如 NumPy、Matplotlib、Scikit-learn) 进行数据分析、可视化与建模。	50%
			树立科技报国与创新驱动发展的使命感, 学生将认识到金融科技对国家经济发展的重要性, 激发科技报国的使命感, 培养创新精神与实践能力和实践能力, 努力将所学知识应用于金融领域的自主创新, 助力我国金融科技的国际化与竞争力提升。	50%
L07	①	M	学生将深刻理解数据安全与隐私保护的重要性, 遵守相关法律法规, 在数据采集、存储与分析过程中恪守职业道德, 杜绝非法爬取、滥用数据等行为。	50%

		树立科技报国与创新驱动发展的使命感，学生将认识到金融科技对国家经济发展的重要性，激发科技报国的使命感，培养创新精神与实践能力，努力将所学知识应用于金融领域的自主创新，助力我国金融科技的国际化与竞争力提升。	50%
		学生将深刻理解数据安全与隐私保护的重要性，遵守相关法律法规，在数据采集、存储与分析过程中恪守职业道德，杜绝非法爬取、滥用数据等行为。	50%
②	M	树立科技报国与创新驱动发展的使命感，学生将认识到金融科技对国家经济发展的重要性，激发科技报国的使命感，培养创新精神与实践能力，努力将所学知识应用于金融领域的自主创新，助力我国金融科技的国际化与竞争力提升。	50%

### 三、课程内容与教学设计

#### (一) 各教学单元预期学习成果与教学内容

<p>第一单元 Python 基础</p> <p>本单元聚焦大数据技术及其应用，帮助学生掌握大数据的基本概念、核心特征及其时代背景，并培养运用工具解决实际问题的能力。课程将简明介绍 Python 基本语法，包括变量、数据类型、控制结构、函数与模块等，并展示其在 Web 开发、数据分析及机器学习等领域的应用。结合课程重点，将深入讲解网络爬虫技术，涵盖 HTTP 协议、网页结构解析、数据提取方法及反爬虫策略，通过实践帮助学生掌握爬虫程序设计与实现，为大数据分析提供数据支持。课程要求学生理解大数据概念与技术框架，掌握 Python 编程并完成数据分析与爬虫任务，通过团队合作完成综合性项目，同时积极参与课堂讨论，培养解决复杂问题的能力。</p> <p>理论课时数：14；实验课时数：0</p>
<p>第二单元 Python 数据分析基础</p> <p>主要工作内容包括负责从各种来源收集数据，并对数据进行预处理和清洗。使用机器学习和深度学习技术来开发模型，并进行调试和优化以提高模型的准确性和性能。不断跟进最新的机器学习和深度学习算法，并进行实验以确定哪种算法最适合解决具体的问题。分析模型的结果，并撰写报告以说明模型的性能和应用场景。与团队成员包括数据科学家、软件开发人员、产品经理等合作，并与非技术人员沟通解释人工智能模型的工作原理和应用场景。</p>

理论课时数：0；实验课时数：20

### 第三单元 Python 数据分析进阶

主要工作内容包括负责从各种来源收集数据，并对数据进行预处理和清洗。使用机器学习和深度学习技术来开发模型，并进行调试和优化以提高模型的准确性和性能。不断跟进最新的机器学习和深度学习算法，并进行实验以确定哪种算法最适合解决具体的问题。分析模型的结果，并撰写报告以说明模型的性能和应用场景。与团队成员包括数据科学家、软件开发人员、产品经理等合作，并与非技术人员沟通解释人工智能模型的工作原理和应用场景。

理论课时数：0；实验课时数：10

### 第四单元 大作业及实验展示

本单元为大作业及实验展示展示

理论课时数：2；实验课时数：2

## (二) 教学单元对课程目标的支撑关系

课程目标 教学单元	1	2	3	4	5	6	7	8	9
	第一单元 Python 基础	√	√	√	√	√	√	√	√
第二单元 Python 数据分析基础	√	√	√	√	√		√	√	
第三单元 Python 数据分析进阶	√	√	√	√			√	√	
第四单元 大作业及实验展示			√			√		√	√

## (三) 课程教学方法与学时分配

教学单元	教与学方式	考核方式	学时分配		
			理论	实践	小计
第一单元 Python 基础	讲授法	大作业	14	0	14
第二单元 Python 数据分析基础	讲授法、直观演示法、讨论法、理实一体化	实验报告	0	20	20
第三单元 Python 数据分析进阶	讲授法、直观演示法、讨论法、理实一体化	实验报告	0	10	10
第四单元 展示与考核	大作业及实验展示	考核	2	2	4
合计			16	32	48

## (四) 课内实验项目与基本要求

序号	实验项目名称	目标要求与主要内容	实验时数	实验类型
1	电影评论数据爬取	数据成为我们获取和传递信息的重要载体。无论是电影评论、新闻报道，还是学术论文、产品评论，社交媒体文本等蕴含着丰富的信息。电影评论数据作为反映观众喜好、电影质量以及市场趋势的重要数据之一，对电影产业、影评人、市场营销人员等都具有极高的价值。数据爬取是获取数据的重要方式之一。学习如何使用 Python 进行电影评论数据的爬取，从网络上抓取有用的数据，以便进行后续的分析 and 应用。	20	③
4	人机交互流程设计	本实验在使用交互流程基础上，给出在图像识别评估系统中使用“resnet.onnx”模型的一种人机交互的最优方式。ResNet 作为一种深度卷积神经网络架构，凭借其深度残差连接机制，能够有效缓解梯度消失问题，实现更深层次的网络结构，从而捕获更加丰富和抽象的图像特征，极大地提高了图像识别的准确性和效率，推动了人工智能技术在现实世界中的广泛应用和商业化进程。	12	③

实验类型：①演示型 ②验证型 ③设计型 ④综合型

#### 四、课程思政教学设计

在课程教学中，将通过理论讲解、案例分析与实践操作相结合的方式，帮助学生深刻理解数据安全与隐私保护的重要性。首先，在理论部分，课程将系统介绍《网络安全法》《数据安全法》《个人信息保护法》等相关法律法规，明确数据采集、存储与分析过程中的法律边界与道德要求。例如，讲解非法爬取数据的法律后果及滥用数据对个人隐私与社会安全的危害。其次，通过案例分析，展示因数据泄露、非法爬取或滥用数据引发的重大事件（如某金融机构因数据泄露导致用户信息被恶意利用），引导学生思考数据安全的重要性及其对社会的影响。最后，在实践环节中，课程将设置数据采集与处理的模拟场景，要求学生严格遵守数据使用规范，设计合法、合规的爬虫程序，并在数据存储与分析过程中采取加密、匿名化等安全措施。通过理论与实践的结合，学生将树立正确的数据使用观念，增强法律意识与职业道德，成为负责任的数据从业者。

同时，通过小组讨论与辩论，引导学生探讨数据使用中的伦理问题（如数据所有权、数据共享与隐私保护的平衡），培养其批判性思维与社会责任感。通过这些教学设计，学生不仅能够掌握数据安全的技术手段，还能够在未来的工作中自觉遵守法律法规，维护数据安全与社会公共利益。

课程将通过融入国家战略、行业前沿与创新实践，帮助学生树立科技报国与创新驱动发展的使命感。首先，在课程导论部分，将结合我国金融科技的发展现状与国际化竞争态势，阐述金融科技对国家经济发展的重要性。例如，介绍我国在移动支付、区块链、人工智能等领域的领先地位，以及金融科技在服务实体经济、推动普惠金融中的作用。通过具体案例（如蚂蚁集团、腾讯金融科技等企业的创新实践），激发学生的民族自豪感与使命感。其次，在课程内容中，将融入创新驱动发展的理念，鼓励学生探索新技术、新方法。例如，在金融数据分析与策略设计环节，引导学生结合我国金融市场特点，设计具有创新性的量化策略；在数据存储与处理环节，鼓励学生研究分布式存储与计算技术，提升数据处理效率。

此外，课程将设置创新实践项目，要求学生以团队形式完成一个金融科技领域的创新课题。通过项目实践，学生不仅能够将所学知识应用于实际问题，还能够培养团队协作能力与创新精神。同时，课程将组织学生参加金融科技领域的创新创业大赛或行业论坛，拓宽视野，了解行业前沿动态，激发其投身金融科技创新的热情。通过这些教学设计，学生将深刻认识到金融科技对国家战略的意义，树立科技报国的远大理想，努力成为推动我国金融科技国际化与竞争力提升的中坚力量。

#### 五、课程考核

总评构成	占比	考核方式	课程目标									合计
			1	2	3	4	5	6	7	8	9	
X1	40%	大作业及实验展示	10	10	10	10	10	10	20	10	10	100

X2	30%	实验报告			10	30	20	30	10			100
X3	30%	课堂表现考勤	20	20	10					20	30	100

评价标准细则 (选填)

考核项目	课程目标	考核要求	评价标准			
			优 100-90	良 89-75	中 74-60	不及格 59-0
1						
X1						
X2						
X3						
X4						
X5						

六、其他需要说明的问题

--