

数据科学导论课前导读

欢迎进入《数据科学导论》的课程学习！在当今数字化时代，数据科学已经成为推动科技进步、经济发展和社会变革的核心力量。从人工智能到大数据，从商业智能到科学研究，数据科学的应用无处不在，而大模型技术的崛起更是为这一领域带来了前所未有的机遇和挑战。

一、数据科学的核心地位

数据科学是一门跨学科的领域，它结合了数学、统计学、计算机科学和领域知识，旨在从海量数据中提取有价值的信息。通过数据预处理、数据挖掘、机器学习和深度学习等技术，数据科学家能够解决复杂的问题，为决策提供支持。无论是金融风险预测、医疗诊断、还是智能交通，数据科学都扮演着不可或缺的角色。

二、大模型技术的崛起

近年来，大模型技术（如人工智能中的 Transformer 架构、GPT 系列等）在数据科学领域迅速崛起，成为推动行业发展的关键力量。以书卷侠（<https://scholarhero.cn/>）为例，其开发背景正是基于大模型的强大语言理解和生成能力。大模型通过海量数据的预训练，能够自动学习语言的语义和语法特性，从而在自然语言处理任务中表现出色。这种技术不仅提高了数据处理的效率，还为数据科学的应用场景带来了更广阔的想象空间。

三、大模型在数据科学中的重要性

1. **提升数据处理效率：**大模型能够处理和分析海量数据，快速提取关键信息，减少人工干预。
2. **增强模型泛化能力：**通过预训练和微调，大模型能够在不同领域和任务中表现出色，适应多样化的应用场景。
3. **推动跨领域应用：**大模型为数据科学在不同领域的融合提供了技术支持，例如在医疗影像分析、智能教育和科学研究中的应用。
4. **降低技术门槛：**大模型的开源和商业化应用，使得更多的企业和研究者能够快速上手，加速数据科学的普及。

四、课程学习目标

教学目标：

1. 介绍数据科学的基本思想、发展历程。
2. 介绍数据科学中常见的分析方法、建模思想、常见算法和基本工具。
3. 结合案例讲解数据分析在实际生活中的应用。
4. 介绍数据科学和传媒行业的结合应用。
5. 探索基于大模型与数据科学结合的应用。

通过本课程的学习，你将：

1. 理解数据科学的基本概念和发展历程。
2. 掌握数据预处理、数据挖掘和机器学习的核心方法。
3. 学习大模型技术的基本原理及其在数据科学中的应用。
4. 培养解决实际问题的能力，通过案例分析和实践项目，将理论知识应用于实际场景。

五、课程内容预览

本课程分为四大部分：

1. **数据思维**：探讨数据科学的起源、数据思维的特点及 DIKW 模型的应用。
2. **数据价值**：涵盖数据预处理、描述性分析和探索性分析。
3. **数据技术**：深入讲解数据存储、计算、挖掘、机器学习和深度学习。
4. **数据未来**：畅想数据科学的未来发展方向。

六、语结

在学习《数据科学导论》的过程中，希望你能深刻体会到数据科学的魅力和大模型技术的强大潜力。通过理论学习与实践操作的结合，你将能够为未来的数据驱动世界贡献自己的力量。让我们一起开启这段充满挑战与机遇的学习之旅吧！

任课老师：王小宁 邮箱：sdwangxiaoning@cuc.edu.cn

办公地点：46-204

计算广告专业：

时间地点：1-9 周, 11-17 周 星期四 7-8 节 四十八教 402

数据科学与大数据技术专业：

时间地点：1-9 周, 11-17 周 星期二 3-4 节 四十八教 707

完成教学的工具（本学期主要使用 R 语言）：

R 软件，需提前在本地电脑下载 R(<https://cran.r-project.org/>) 和 Rstudio(<https://rstudio.com/>)

注：Windows 系统或 Mac 系统下载对应的系统版本即可。

Typora：完成报告撰写。

注册课程助教平台书卷侠：<https://chat.scholarhero.cn/>

加入数据科学课程，邀请码：349275

考核：

平时作业：30%，课堂表现，布置 4-5 次课程作业

期中测试：40%

期末测试：30%，课程报告

注：本课程无固定参考书，但有大量的阅读资料和相关的资源推荐，会以电子版给大家。

前两周学习任务及材料：

第一周：

阅读图书：《统计与真理，怎样运用偶然性》，C. R. 劳著，科学出版社，2004

电影：The Man Who Knew Infinity（知无涯者），2015

第二周:

材料: 十字路口的统计学: 谁在应对挑战? (中文, 2020)

<https://cosx.org/2020/01/crossroad-statistics-who-meet-challenge-1/>

<https://cosx.org/2020/01/crossroad-statistics-who-meet-challenge-2/>

<https://cosx.org/2020/01/crossroad-statistics-who-meet-challenge-3/>

学数据科学与大数据技术专业, 我们最应该学些什么?

<https://mp.weixin.qq.com/s/S2UUyhntTCHtunEzdyf1g>

数据科学家能力素质模型

<https://mp.weixin.qq.com/s/ls5JuoX7wWvZJcJLraYcbA>

数据科学 50 年: https://mp.weixin.qq.com/s/s6dxiCMLhXrDc3pJJ_jupg

历年数据科学导论优秀结课报告:

合集|数据科学导论课程报告展示-2022

<https://mp.weixin.qq.com/s/Bx44Q8GzRDstYioOoqPA5Q>

合集|数据科学导论课程 2024 春季优秀报告展示

<https://mp.weixin.qq.com/s/DxO-OvoTLsd2CQSiVVpHHg>

《统计与真理—怎样运用偶然性》读后感

https://mp.weixin.qq.com/s/lds0Dz14r0Mwgzv_tTEO8Q

《知无涯者》观后感

<https://mp.weixin.qq.com/s/Oboq0R3LTfQ2ZlciTaVL1Q>

电影: A Beautiful Mind (美丽心灵), 2001

第一次作业:

选择前两周的阅读材料或电影写一篇读后感, 文字 1500-2000 字, 提交时间 2025 年 3 月 13 日前, 提交畅课平台。