

《数据挖掘与商务分析》教学大纲

(2025—2026 年第 1 学期)

课程: 数据挖掘与商务分析

课程序号: 0518

课程代码: 100765

授课教师: 肖升生

答疑时间: 在线答疑+线下预约

办公室: 科研实验大楼 837 室

E-mail: xiao.shengsheng@shufe.edu.cn

电话: 021-65904410-837

课程类别: 通识模块六（科技进步与科学精神）限选课

课程安排说明: 2025 年 9 月 8 日—2025 年 12 月 26 日

上课时间: 周一 15:25—16:10

授课地点: 科研实验大楼 603, 国定路校区

授课形式: 线下教学

教学学时分配表:

学分	总学时	理论教学学时	实践教学学时	其他
2	32	23	9	0

课件网址: <https://canvas.shufe.edu.cn/courses/35518>

视频网址: <https://www.xueyinonline.com/detail/215844957>

教材和参考书目:

指定教材:

《商务智能方法与应用（第 2 版）》，刘红岩，清华大学出版社， ISBN: 9787302558101

参考教材:

《数据挖掘导论（完整版）》，Pang-Ning Tan，人民邮电出版社， ISBN: 9787115241009

预备知识:

常见概率分布、期望方差、微积分、梯度下降等优化方法、编程基础

先修课程: 概率论与数理统计、高等数学、编程基础

课程达成目标:

本课程旨在引导学生系统掌握数据挖掘在商业管理领域中的基本原理、主流技术与典型

应用场景，重点培养学生以数据为基础发现问题、分析问题并提出解决方案的综合能力。通过对分类、聚类、关联分析、异常检测、预测建模等典型算法的讲解与实践训练，学生将具备使用 Python 等主流数据分析语言和工具独立完成数据处理、模型构建与结果解释的能力。课程特别注重将数据挖掘方法与营销、运营、金融等管理实际问题结合，提升学生将算法转化为商业洞察的实践能力。

此外，课程在教学过程中嵌入思政育人元素，注重学生技术伦理意识、数据隐私保护意识和社会责任感的培养。通过事件分析与技术反思，引导学生理性认知数据技术的社会影响，树立服务社会、服务国家的价值导向，提升信息素养和创新精神，最终实现专业能力与价值观塑造的双重提升。

课程设置知识要求：

《数据挖掘与商务分析》是一门融合数据科学与管理决策的交叉课程，对学习者在数学基础、编程技能与管理素养等方面提出了多维度的知识准备要求。首先，学生应具备扎实的概率论与数理统计知识，包括随机变量、概率分布、统计推断、线性回归等，为理解数据挖掘模型的构建与评估提供基础支撑。其次，应掌握一定程度的高等数学知识，特别是函数导数、偏导数与优化原理，这对于深入理解模型的训练机制（如梯度下降）具有关键意义。此外，学生需具备编程基础，能够使用 Python 或 R 进行数据读取、清洗、可视化和基本建模操作，掌握如 pandas、numpy、ggplot2、matplotlib 等数据分析常用工具包的使用方法。在人文与管理素养方面，学生应具备初步的商业逻辑与管理问题意识，能理解诸如客户关系管理、库存优化、用户推荐等应用背景。同时，建议具备一定的信息伦理与法律意识，理解数据使用中的隐私保护、算法公平性等问题。

通过这些前置知识的支撑，学生能够在本课程中更有效地理解数据挖掘技术的本质、掌握其在商业管理中的应用逻辑，并为后续课程（如《商业智能》、《人工智能》等）奠定坚实基础。

课程设置能力要求：

本课程通过“理论讲授 + 案例分析 + 编程实操 + 项目实践”的教学方式，系统培养学生以数据为核心分析和解决商业管理问题的综合能力。首先，课程注重培养学生的数据思维与问题建模能力。通过对管理实际问题的抽象、变量识别与建模过程训练，学生能够将商业管理情境转化为可量化、可分析的问题形式，形成数据驱动决策的逻辑思维方式。其次，在实际操作层面，课程强化学生的数据处理与算法实现能力。通过使用 Python 语言，学生将在真实或模拟的商业数据集上开展预处理、建模与可视化练习，掌握分类、聚类、推荐、关联规则等常用模型的编程实现，具备独立完成小型数据挖掘项目的能力。在此基础上，课程进一步提升学生的跨领域分析与管理解释能力。通过案例分析与结果呈现训练，提升其“从数据中讲出故事”的表达力和策略建议能力。此外，课程特别设置阶段性项目、小组协作任务与汇报环节，系统锻炼学生的团队协作、问题拆解、逻辑组织与表达汇报能力，增强其在复杂环境中开展数据分析工作的实战素养。

“课前预习”机制作为课程常规要求，学生可通过观看提前录制好的视频（链接：<https://www.xueyinonline.com/detail/215844957>），提前掌握背景知识和提出问题，为课堂讨论和实操做准备。

考核形式：

将各章节内容适时举行课堂讨论、随堂测验，安排个人平时作业和 1 次期末团队项目报告。其中，团队作业以不超过 3 人一组协作完成并需要进行项目汇报，锻炼学生运用所学

知识解决领域问题的能力。课程最终考核方法如下：

课堂参与	10%
随堂测验	20%
平时作业	30%
期末团队实践项目报告（具体要求另附）	40%

学术诚实

涉及学生的学术不诚实问题主要包括考试作弊；抄袭；伪造或不当使用在校学习成绩；未经老师允许获取、利用考试材料。对于学术不诚实的最低惩罚是考试给予 0 分。其它的惩罚包括报告学校相关部门并按照有关规定进行处理。

课程具体教学要点和设计

一、课程导论

本模块帮助学生构建对数据挖掘、人工智能与商务智能融合关系的初步认识，厘清数据在商业场景中所扮演的价值角色，并从技术和业务双重视角理解数据驱动决策的重要性。通过介绍数据类型、数据挖掘发展脉络与课程结构安排，引导学生形成整体学习框架，同时结合数据隐私保护法规与跨行业的数据应用流程，引导学生树立负责任的技术意识。

- **教学要点：**数据类型与价值识别；数据挖掘与 AI 大模型关系；数据隐私保护意识；跨行业数据分析流程；课程模块学习路径。
- **重点：**数据挖掘与 AI/BI 融合逻辑；数据赋能商业问题的整体链条。
- **难点：**学生初次接触跨领域内容，需在技术概念与商业应用之间建立清晰连接。

主要授课内容及知识点安排：

1. 数据类型与价值使用
2. 数据挖掘、AI 大模型与商业智能
数据挖掘与隐私保护
3. 跨行业的数据挖掘流程
4. 课程内容与设计
5. 课程学习材料

二、数据预处理

数据预处理是数据挖掘实践中的第一步，是决定模型质量与分析可信度的关键环节。本部分讲授从原始数据采集到清洗、规约、变换的全流程操作，强调对缺失值、异常值、多重变量冗余的处理方法，同时引导学生注意数据获取与使用过程中的伦理与合规问题。通过典型案例训练学生数据整合与质量管理能力，为后续建模奠定基础。

- **教学要点：**数据清洗、抽样、规约、变换方法与技术伦理；预处理流程与案例。
- **重点：**数据质量与预处理方法的匹配；对业务目标的支持性判断。
- **难点：**高维数据的规约选择；复杂数据源的清洗与整合策略搭建。

主要授课内容及知识点安排：

1. 数据预处理概述
数据获取的技术伦理
2. 数据清理
3. 数据抽样
4. 数据规约
5. 数据变换
6. 数据预处理案例

三、分类分析与商务实践

分类分析是监督学习的核心内容之一，广泛应用于客户信用评估、市场响应预测等领域。本模块系统讲授决策树、朴素贝叶斯、逻辑回归等典型分类方法，辅以实际业务背景中的案例分析，引导学生理解模型假设、构建过程及评估指标如准确率、AUC 等。通过多模型比较与可解释性分析，训练学生在实际商业环境中做出建模选择并输出可执行结论的能力。

- **教学要点：**分类问题定义；常用分类模型及其适用场景；分类性能评估指标；真实业务案例分析。
- **重点：**模型选择与性能评估机制；分类模型在商业中的可解释性与业务含义。
- **难点：**处理样本不平衡；理解复杂模型参数与调优技巧对实际效果的影响。

主要授课内容及知识点安排：

1. 分类分析基本概念
2. 决策树分类模型
3. 朴素贝叶斯模型
4. 逻辑回归模型
5. 分类模型评估
6. 商务案例分析

四、 关联规则分析与商务实践

本部分聚焦于揭示数据项之间的潜在共现关系，适用于购物篮分析、产品推荐与营销策略优化。课程将介绍频繁项集挖掘算法如 Apriori 的基本原理，讲解支持度、置信度、Lift 等核心指标的使用方式，同时强调如何通过指标筛选出“可落地”的业务洞察，规避冗余或无效规则。

- **教学要点：**关联规则的数学定义与应用逻辑；频繁项集与规则生成算法；典型商业应用案例。
- **重点：**度量指标解释与合理使用；对业务含义强的规则筛选。
- **难点：**高维稀疏数据下的规则挖掘；提升从数据中“挖出价值”的能力。

主要授课内容及知识点安排：

1. 关联规则分析基本概念
2. 频繁项集生成
3. 关联规则生成
4. 关联规则分析拓展
5. 商务案例分析

五、 聚类分析与商务实践

聚类分析作为非监督学习的重要方法，在客户细分、产品定位、市场划分等应用中极具价值。本模块系统介绍 K-means、层次聚类等方法及其数学基础，引导学生理解聚类流程、相似性度量方式及聚类数选择策略，并通过案例探讨聚类结果与营销、产品策略的融合方式。

- **教学要点：**聚类方法与流程；聚类数确定与模型评估；客户分群等实践案例。
- **重点：**聚类算法与业务目标的匹配设计；聚类结果的可解释性与实际转化能力。
- **难点：**数据聚类效果分析；高维空间中的距离度量选择与理解。

主要授课内容及知识点安排:

1. 聚类分析基本概念
2. 聚类分析基本流程
3. 聚类分析主要方法
4. 聚类分析效果评估
5. 商务案例分析

六、 自然语言处理与商务实践

本部分聚焦于文本数据挖掘与自然语言处理（NLP）技术在商业中的应用，包括客户评论分析、舆情追踪、智能客服等场景。通过介绍 TF-IDF、主题模型、词向量及大语言模型（如 GPT）的基本原理，帮助学生理解文本结构化处理与模型建构方法。

- **教学要点：**文本挖掘方法；词向量与主题建模；大语言模型基础与案例。
- **重点：**文本表示方法与建模手段的结合应用；LLM 在商业中的实际落地。
- **难点：**非结构化文本清洗处理；模型输出可信度评估。

主要授课内容及知识点安排:

1. 文本挖掘、自然语言处理（NLP）基本概念
2. 文本数据预处理
3. 文本表示-向量空间模型
4. 文本表示-主题模型&词向量模型
5. 大语言模型
大语言模型中的伦理问题
6. 商务案例分析

七、 时空数据挖掘与商务实践

本模块面向时序与地理数据的挖掘与可视化分析，讲授时间序列建模、空间热力图分析及轨迹聚类等方法，应用于零售选址、城市人流预测等商业场景。通过实际案例提升学生对时空多维数据的综合处理与表达能力。

- **教学要点：**时空数据结构；可视化工具与动态展示方式；典型商业场景应用。
- **重点：**时间/空间特征的建模差异性；跨时间点、跨区域的关联分析。
- **难点：**多维异质数据整合；空间分析中边界效应与尺度误差控制。

主要授课内容及知识点安排:

1. 时空数据挖掘应用场景及基本概念
2. 时空数据管理
3. 时空数据可视化
4. 时空数据挖掘算法
5. 商务案例分析

八、 社交网络分析与商务实践

本部分围绕用户之间的互动与传播关系建模展开，涵盖图结构基础、中心性指标、社群识别算法及影响力扩散模型。通过 KOL 识别、产品扩散路径分析等举例，训练学生理解关系型数据如何影响商业策略。

• **教学要点：**图数据结构与社交网络基础；节点分析与社群发现；网络影响力度量与路径建模。

• **重点：**结构关系识别与商业洞察提炼的结合；复杂网络的关键节点识别。

• **难点：**图计算的算法复杂性；多层网络融合建模与解释。

主要授课内容及知识点安排：

1. 社交网络分析基本概念
2. 图论基础
3. 社群识别
4. 影响力度量
5. 商务案例分析

九、 推荐系统与商务实践

推荐系统是商业数据挖掘的核心应用领域之一，涉及内容推荐、协同过滤、矩阵分解等多种技术方法。学生将在理解推荐算法逻辑的基础上，掌握用户画像构建与冷启动应对策略，并通过案例分析强化推荐结果的精准性与可控性。

• **教学要点：**推荐系统架构；主流推荐算法原理；实际商业平台的应用场景。

• **重点：**算法选择与用户行为建模的协同机制；推荐效果优化策略。

• **难点：**冷启动问题处理；多源行为数据的建模与融合。

主要授课内容及知识点安排：

1. 推荐系统基本概念
2. 基于内容的推荐
3. 基于近邻的协同过滤
4. 基于矩阵分解的协同过滤
5. 商务案例分析

教学具体安排

序号	内 容	课时数	进度
1	课程导论	2	第 1 周
2	数据预处理	4	第 2、3 周
3	分类分析与商务实践	4	第 4、5 周
4	关联分析与商务实践	2	第 6 周
5	聚类分析与商务实践	4	第 7、8 周
6	自然语言处理与商务实践 随堂测验	6	第 9、10、11 周
7	时空数据挖掘与商务实践	2	第 12 周
8	社交网络分析与商务实践	2	第 13 周
9	推荐系统与商务实践	4	第 14、15 周
10	商务智能实践项目/案例报告	2	第 16 周
合计		32	

注：第 16 周为学生进行期末团队项目课堂展示