《数据挖掘方法与技术》教学大纲

课程编号：2121534A

课程类型：□通识教育必修课 □通识教育选修课

□学科基础课 ☑专业核心课

□专业提升课 □专业拓展课

总 学 时：64 讲课学时：32 实验（上机）学时：32

学　　分：4

考试类型：☑考试 □考查

适用对象：信息管理与信息系统专业

□是 ☑否 适合作为其他专业学生的个性化选修课

先修课程：概率论与数理统计、高等数学、数据分析理论与实践（双语）

一、教学目标

本课程主要介绍数据挖掘的理论知识、主流技术和实战应用，使学生通过对实际数据的挖掘，更加深入地理解常用的数据挖掘算法和模型。培养学生的数据挖掘的理论分析与应用实践的综合能力，学生能够从具体应用场景出发，利用所学的知识解决行业应用问题。

本课程具有很强的理论性与实践性，目标是使学生通过本课程所规定的全部教学内容的学习，能够对数据挖掘概念、分类、聚类、关联规则等内容有基本的认识，熟悉数据挖掘的基本算法、工具以及掌握其技术原理和应用方法，了解数据挖掘方向的主要研究热点与应用领域，为进入更深入的智能数据分析研究打好基础。

在本课程中，将伦理道德和信息安全等问题作为课程思政的载体，不仅向学生解释相关技术的原理和概念，而且注重在知识传播中强调价值引领，使学生在运用数据挖掘方法与技术解决问题过程中树立正确的伦理道德观念和法制观念。

二、教学内容及其与毕业要求的对应关系

本课程系统讲授数据挖掘的方法和技术，包括数据挖掘基本概念、分类、聚类、关联规则、数据挖掘应用案例等内容，并深度剖析了数据挖掘中的基本概念及各种算法的原理，对数据挖掘的数据预处理、分类、聚类和关联规则等方法的具体实现进行了详细的讲述。

课程注重理论与实践相结合，采取实例教学法、小组教学法等多种教学方法进行授课。要求学生在实验环节掌握数据挖掘算法的基本原理、思路及在决策中的实战应用。要求学生课后完成基于课堂和实验教学内容的扩展实践练习，进一步熟悉挖掘技术和算法、深化相关领域的实际应用场景练习。

通过本课程的学习从以下方面促进了毕业要求的实现：熟练掌握数据挖掘领域的核心技术，并能运用这些理论、知识、原理和方法解决数据挖掘相关的问题；掌握人工智能大环境下进行数据挖掘和处理的能力，具有较宽的知识面和知识结构；具有自主学习和终身学习意识，有创业创新能力及不断学习与适应社会发展与改革创新的能力。

三、各教学环节学时分配

以表格方式表现各章节的学时分配，表格如下：

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 章节内容 | 讲课 | 实验 | 其他 | 合计 |
| 第一章 | 数据挖掘概述 | **2** | **2** |  | **4** |
| 第二章 | 认识数据与预处理 | **2** | **2** |  | **4** |
| 第三章 | 关联规则挖掘 | **6** | **6** |  | **12** |
| 第四章 | 聚类 | **8** | **8** |  | **16** |
| 第五章 | 分类 | **9** | **9** |  | **18** |
| 第六章 | 数据挖掘应用案例 | **5** | **5** |  | **10** |
| **合计** |  | **32** | **32** |  | **64** |

四、教学内容

第一章 数据挖掘概述

教学内容：介绍数据挖掘的基本概念、数据挖掘过程、数据挖掘的主要任务以及数据挖掘使用的主要技术，数据挖掘的应用和面临的问题，对数据挖掘能够解决的问题和解决问题思路有清晰的认识。

课程思政方面，通过讲授数据挖掘的实际应用, 比如利用Facebook个人信息分析来影响美国大选的案例入手，引发学生对个人信息泄露和数据挖掘用途是否合乎伦理要求的思考，引导学生思考数据挖掘涉及到的伦理道德问题。

重点和难点：数据挖掘的概念与主要任务。

考核要求：了解数据挖掘的定义和作用，掌握数据分析与数据挖掘的联系，理解和掌握数据挖掘的主要内容，理解数据挖掘中的主要问题。

实验一、数据分析与可视化实训

第二章 认识数据与预处理

教学内容：数据的属性及其类型、基本统计描述，数据对象的相似性度量，数据清理及Python数据清洗方法、数据集成、数据变换与数据离散化、数据归约。

课程思政方面讲授数据来源的伦理风险, 比如通过Cookie数据收集、APP权限和用途不对等案例说明数据来源的伦理风险，讨论哪些是合法和非法的数据获取途径。

重点和难点：数据对象的属性及其分类，数据对象的相似性度量，数据清洗内容， 数据集成过程中的关键问题，数据变换和收据归约方法。

考核要求：了解数据的属性类型，理解数据的基本统计描述, 掌握度量数据相似性和相异性的方法；理解数据清理、数据集成、数据规约、数据变换于数据离散化的方法, 掌握数据预处理的基本方法。

实验二、利用Python实现数据合并、数据清洗、数据变换

第三章 关联规则挖掘

教学内容：频繁项集、闭项集和关联规则，Apriori算法，频繁模式增长树算法，使用垂直数据格式挖掘频繁项集，关联模式评估方法。

课程思政方面，通过实例讲授数据挖掘获取商业信息的过程中客户隐私保护的问题，引导学生思考如何确保数据不被滥用。

重点和难点：Apriori算法，FP频繁模式增长树算法，关联规则的评估方法。

考核要求： 了解频繁项集、闭项集和关联规则的概念，理解模式评估方法, 掌握Apriori算法和挖掘频繁项集的模式增长方法。

实验三、Python实现数据的关联规则分析

第四章 聚类

教学内容：聚类算法分类，K-Means算法，层次聚类，基于密度的聚类。

课程思政方面，通过美国政府用人工智能“把守国门”的极端审查计划的案例，和欧盟API-PNR系统中的数据使用的相称性原则案例，说明政府在应用数据时要注意目的的正当和方法的合理性问题。

重点和难点：聚类算法的原理，层次聚类算法，基于密度的聚类算法思想。

考核要求：了解聚类的概念，掌握k-Means 和k-Medoids算法、层次方法和基于密度的方法等典型算法及其Python的实现方法。

实验四、数据的聚类分析综合实验

第五章 分类

教学内容：分类概述，决策树原理，ID3算法，C4.5算法，CART算法，树剪枝，K-近邻算法，SVM算法，贝叶斯分类，模型评估与选择，组合分类，BP神经网络。

重点和难点：分类的原理，决策树规约算法，KNN算法，SVM算法，贝叶斯分类方法，Adaboost，随机森林，神经网络的原理。

课程思政方面,通过实际应用引导学生认识数据挖掘中的欺诈、信用问题。

考核要求：了解分类的概念，理解评估分类器性能的度量方法, 掌握决策树分类算法、SVM、贝叶斯分类算法、模型评估与选择、组合分类及利用Python实现分类的方法，掌握感知机模型和BP神经网络。

实验五、分类分析综合实验

第六章 数据挖掘应用案例

教学内容：数据挖掘应用案例。

课程思政方面，通过数据挖掘应用案例，让学生从自身专业视角思考数据挖掘获取信息时面临的伦理道德问题。

教学重点、难点：数据挖掘综合应用。

考核要求：掌握利用Python进行综合数据挖掘。

实验六、 数据挖掘案例分析

五、考核方式、成绩评定

本课程采用平时、期末开卷考试相结合的考核方式，尤其注重实践环节的考核。各部分所占比重为平时成绩占50%，期末考试成绩占50%。

平时成绩包括考勤成绩（10%）和作业成绩（40%）。

六、主要参考书及其他内容

[1] 魏伟一，[张国治](https://book.jd.com/writer/%E5%BC%A0%E5%9B%BD%E6%B2%BB_1.html)著，《python数据挖掘与机器学习》，北京：清华大学出版社．清华大学出版社，2021.4

[2] [孙家泽](https://book.jd.com/writer/%E5%AD%99%E5%AE%B6%E6%B3%BD_1.html)，[王曙燕](https://book.jd.com/writer/%E7%8E%8B%E6%9B%99%E7%87%95_1.html)著,数据挖掘算法与应用.北京：清华大学出版社．2020.10

[3] [刘鹏](https://book.jd.com/writer/%E5%88%98%E9%B9%8F_1.html),[张燕](https://book.jd.com/writer/%E5%BC%A0%E7%87%95_1.html),[陶建辉](https://book.jd.com/writer/%E9%99%B6%E5%BB%BA%E8%BE%89_1.html),[姜才康](https://book.jd.com/writer/%E5%A7%9C%E6%89%8D%E5%BA%B7_1.html)著,数据挖掘基础.北京：清华大学出版社．2018.7

[4] [方巍](https://book.jd.com/writer/%E6%96%B9%E5%B7%8D_1.html)著,Python数据挖掘与机器学习实战 北京：机械工业出版社．2019.5

执笔人：曹娜 教研室主任：胡磊　　　　　系教学主任审核签名：