**《数据处理及挖掘》课程教学大纲**

执笔人：荀径 编写日期：2022年11月

**一、课程基本信息**

1．课程编号： M401104B

2．课程平台：专业教育平台

3．课程模块：专业拓展选修课程

4. 课程性质：选修

5. 学时/学分：32学时/2学分

6．先修课程：计算机技术、概率论、数理统计

7．适用专业：信息工程专业

8．教学单位名称：电子信息工程学院

**二、课程目标及学生应达到的能力**

《数据处理及挖掘》是自动化专业本科生的选修课程，着重培养学生对实际工程中的数据科学问题进行描述、抽象建模和解释分析的能力。

本课程的主要任务是通过课堂教学、专题研讨等环节培养学生的创新意识以及数据挖掘理论知识的应用能力，使学生理解数据挖掘相关的基本概念和基本原理，掌握数据挖掘的基本步骤，掌握数据预处理的基本方法，掌握数据挖掘相关软件的使用方法。本课程注重以数据挖掘的发展历史为载体，将专业知识与思政育人元素相融合，引导学生树立正确的人生观、价值观和世界观。

课程教学目标及能力要求具体如下：

1. 掌握数据挖掘分析需要的数学基础，能运用概率论和数理统计等数学知识及计算机系统分析方法，进行数据的清洗和关联性分析。

2．了解数据挖掘技术的作用与发展，掌握数据挖掘的基本概念和一般步骤，对实际问题进行合理建模并利用数据挖掘方法解决实际问题。

3. 通过研究性教学环节，锻炼学生的沟通学习能力，培养学生与老师、合作者进行有效沟通和交流，主动扩展专业外语，追踪专业领域技术发展的前沿，培养独立思考和追踪技术发展趋势的意识。

**三、课程教学目标和毕业要求的对应关系**

|  |  |
| --- | --- |
| 毕业要求指标点 | 课程目标 |
| 5.1 针对工程问题，收集信息、查阅文献、分析现有技术的特点与局限性。 | 1 |
| 9.1 具备人文社会科学素养，理解应担负的社会责任，愿意为社会服务。 | 2 |
| 13.1 具备主动学习的能力，能够运用信息和文献工具，自主学习知识。 | 3 |

**四、课程思政育人目标**

介绍信息时代发展史和中外数据挖掘领域的技术发展方向，坚定新时代的理想信念、厚植爱国主义情怀，深刻理解科技基础能力建设的重大意义。

通过介绍我国科学家的成就和贡献，培养学生坚忍不拔探索的奋斗精神和开拓精神，严谨求实、不迷信不盲从的科学精神，深刻理解科技自立自强的重大意义。

**五、课程教学内容和要求**

总学时 32学时，讲课32学时。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 知识单元 | 知识点 | 要求 | 推荐学时 | 教学方式 | 支撑课程目标 |
| 1 | 绪论 | 数据挖掘的概念 | 掌握 | 2 | 讲授 | 1 |
| 数据挖掘的功能 | 掌握 | 讲授 |
| 数据、信息、知识的区别 | 了解 | 讲授 |
| 数据挖掘的基本工作原理 | 掌握 | 2 | 讲授 |
| 数据挖掘发展简史及发展方向 | 了解 | 讲授 |
| 2 | 数据 | 数据类型 | 掌握 | 2 | 讲授 | 1 |
| 数据质量 | 掌握 | 讲授 |
| 数据预处理 | 掌握 | 2 | 讲授 |
| 相似性与相异性的度量 | 掌握 | 2 | 讲授 |
| 3 | 探索数据 | 鸢尾花数据集 | 理解 | 2 | 讲授 | 2，3 |
| 汇总统计 | 掌握 | 讲授 |
| 可视化 | 掌握 | 2 | 讲授 |
| OLAP与多维数据分析 | 理解 | 讲授 |
| 4 | 数据分类（I） | 基本概念 | 理解 | 2 | 讲授 | 2，3 |
| 解决分类问题的一般方法 | 掌握 | 讲授 |
| 决策树归纳 | 理解 | 2 | 讲授 |
| 模型的过分拟合 | 掌握 | 讲授 |
| 评估分类器的性能 | 掌握 | 2 | 讲授 |
| 比较分类器的方法 | 掌握 | 2 | 讲授 |
| 5 | 数据分类（II） | 基于规则的分类器 | 掌握 | 2 | 讲授 | 2，3 |
| 最邻近分类器 | 掌握 | 讲授 |
| 贝叶斯分类器 | 掌握 | 2 | 讲授 |
| 人工神经网络 | 了解 | 2 | 讲授 |
| 支持向量机 | 了解 | 2 | 讲授 |
| 6 | 关联分析 | 基本概念 | 掌握 | 2 | 讲授 | 2，3 |
| 频繁项集的产生 | 掌握 | 讲授 |
| 规则的产生 | 掌握 | 2 | 讲授 |
| 频繁项集的紧凑表示 | 理解 | 2 | 讲授 |
| 产生频繁项集的其他方法 | 理解 | 2 | 讲授 |
| 关联模式评估 | 理解 | 2 | 讲授 |
| 7 | 聚类分析 | 基本概念 | 理解 | 2 | 讲授 | 2，3 |
| K均值方法 | 掌握 | 讲授 |
| 凝聚层次聚类 | 掌握 | 2 | 讲授 |
| 簇评估 | 理解 | 讲授 |
| 8 | 异常检测 | 预备知识 | 掌握 | 2 | 讲授 | 2，3 |
| 统计方法 | 掌握 |  | 讲授 |
| 基于邻近度的离群点检测 | 掌握 | 2 | 讲授 |
| 基于密度的离群点检测 | 掌握 | 2 | 讲授 |
| 基于聚类的检测 | 理解 | 2 | 讲授 |

**五、课程教学安排**

本课是自动化专业的专业选修课，需要运用先修的概率论、数理统计专业数学知识和计算机系统分析方法，通过课堂理论教学讲解基础原理和分析方法，指导课后学生自学部分拓展知识并开展专题研讨。

教学方法以课堂讲授为主。课堂讲授包括教学方法和手段设计；作业方面主要写明布置习题达到的目的与要求，题量、题型要求等，提倡按照教学的层次设计作业，加大专题研究和研究性、综合性大作业的比例，可限定比例要求。

**1．课堂讲授**

在教学过程中，教师应以建立概念、形成整体思路为基础，指出每章的重点和难点部分。上课时，适当提出一些问题，引导学生分析讨论，以调动学生的主动性，活跃课堂气氛提高课堂质量，集中学生注意力，培养学生发现问题和分析问题的能力。对教学媒体的运用须密切结合课程知识点的需要加以选择。

**2．习题及指导分析**

本课程理论性强，知识点多，需要足够的作业支撑。题型包括计算、画图、分析和设计等。通过习题巩固知识点的掌握，并适当布置自学作业，达到培养学生主动学习拓展知识的能力，锻炼其分析问题解决问题的能力。根据情况适当增设课堂习题指导及作业评述。

**3．专题研究讨论**

通过专题研讨活动，培养学生主动学习意识和举一反三和探索质疑的思维方式，锻炼学生自主学习能力。结合教学内容进展设置适当的专题，可以包括仿真分析、理论分析、新技术追踪等，布置学生以个人或分组方式完成理论或仿真分析、撰写分析报告、课堂报告分析结果并开展讨论。

研讨过程组织的参考形式：学生个人或分组完成数据分析后，提交书面分析报告，并进行课堂口头报告和讨论，对关键问题进行评讲。在课堂讨论中，引导同学交流研究心得，提出疑难问题，开展讨论，对分析研究的结果进行评价比较，从而使学生能够相互借鉴，从不同角度获得对同一问题的充分理解，提出对同一问题的多个解决方案，相互促进，共同提高。

**4. 课程思政**

本课程主要介绍算法实现特点，在坚持正确的中国特色社会主义思想的指导下，围绕立德树人的根本宗旨，将课程思政典型案例（如下表所示）有机融入相关知识点和教学环节中，通过案例讲解、学生提问、课后调研、课堂展示等方式，在进行知识传授、能力培养的同时，润物无声地对学生进行价值塑造。

**表1《数据处理及挖掘》思政案例**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 章节名称 | 知识点 | 所挖掘的思政元素 |
| 1 | 绪论 | 数据科学发展历程 | **【文化自信、科学精神、创新精神】**  **通过阐释“智能”的定义、分析其词源，激发学生对科学探索的兴趣，培养学生严谨认真的学习态度和敢于创新发现的探究精神**  **【方 法】讲授式思政**  **【内 容】**  **1. 教师首先解释智能的定义：《新华词典》: 智慧和能力；**  **2. 探索智能一词的来源，引用《荀子·正名篇》：“所以知之在人者谓之知，知有所合谓之智。所以能之在人者谓之能，能有所合谓之能”。随后请学生朗读此句，并引导学生解读本句。即人所固有的认识外界客观事物的本能叫做知，这种本能与客观万物相合就叫智慧。人固有的掌握外物的能力叫做本能，本能与外物相合的叫做才能。荀子从一般的知识论出发，将“智”解释为人的认识本能和外在事物的相合，即作为一种知识和认识成果的“智”。这种“智”来源于人与物之合，是一个不断积累的过程；**  **3. 最后指出，自然界“四大奥秘”是智能发生、物质本质、宇宙起源、生命本质。本节所讲的内容对于科学探索的积极意义，使学衡学生在听讲、讨论、诵读的过程中，激发对科学探索的兴趣和创新发现的探究精神。** |

**六、课程的考核**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **课程成绩构成及比例** | **考核环节** | | **目标分值** | **考核/评价细则** | **对应的课程目标** |
| 平时成绩100分  占总评成绩的50% | 作业 | 基本概念  基本规律 | 50 | （1）主要考核学生对每章节知识点的复习、理解和掌握程度；  （2）每次作业按30分制单独评分，取各次成绩的平均值作为此环节的最终成绩。 | 1 |
| 基本分析  基本特性 | 50 | 2 |
| 期末考试100分  占总评成绩的50% | 期末  考试 | 基本概念  基本规律 | 50 | （1）采用笔试（闭卷）形式，卷面成绩100分，以卷面成绩乘以其在总评成绩中所占的比例计入课程总评成绩。  （2）考核内容需考察综合运用基本概念、基本原理和分析方法解决数据分析问题的能力，内容需覆盖本课程在教学内容和要求中给出的所有理论知识点，并注意考察各知识点和方法的综合交叉运用。 | 1 |
| 基本分析  基本特性 | 50 | 2 |

**七、本课程与其它课程的联系与分工**

先修课程：概率论与数理统计（B），计算机技术

本课程的先修课程为“概率论与数理统计（B）”和“计算机技术”。需要的概率统计知识主要包括[概率论](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E8%AE%BA)的基本概念、[随机变量](https://baike.baidu.com/item/%E9%9A%8F%E6%9C%BA%E5%8F%98%E9%87%8F)及其[概率分布](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E5%88%86%E5%B8%83)、数字特征、[大数定律](https://baike.baidu.com/item/%E5%A4%A7%E6%95%B0%E5%AE%9A%E5%BE%8B)与[中心极限定理](https://baike.baidu.com/item/%E4%B8%AD%E5%BF%83%E6%9E%81%E9%99%90%E5%AE%9A%E7%90%86)、统计量及其概率分布、[参数估计](https://baike.baidu.com/item/%E5%8F%82%E6%95%B0%E4%BC%B0%E8%AE%A1)和[假设检验](https://baike.baidu.com/item/%E5%81%87%E8%AE%BE%E6%A3%80%E9%AA%8C)、回归分析、[方差分析](https://baike.baidu.com/item/%E6%96%B9%E5%B7%AE%E5%88%86%E6%9E%90)、[马尔科夫链](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E7%A7%91%E5%A4%AB%E9%93%BE)，需要的计算机知识为计算机操作基础知识和数据结构基础知识。

**八、建议教材及教学参考书**

建议教材：

1. Pang-Ning Tan，Michael Steinbach，Vipin Kumar，《数据挖掘导论》，北京，人民邮电出版社，2006年。
2. Peter Harrington，《机器学习实战》，北京，人民邮电出版社，2013年。

教学参考书：

1. 谭建豪等编著，《数据挖掘技术》，北京，中国水利水电出版社，2009年。