**数据挖掘方法与应用课程授课提纲（含课程思政）**

**学年第 学期**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 课程信息 | 课程名称 | 数据挖掘方法与应用 | | | | |
| 开课班级 |  | | | | |
| 课程类型[[1]](#footnote-1) | 专业核心课 | 课程性质[[2]](#footnote-2) | | 含课内实验的理论课 | |
| 学分 | 3 | 周学时 | 3 | 总学时 | 48 |
| 教学周 | 01— 16 周 | 教室 |  | 上课时间 |  |
| 教学安排 | **课堂讲授 11 次，上机实验5 次** | | | | |
| 教师 | 主讲教师 |  | | 电子信箱 |  | |
| *答疑时间* |  | | *答疑地点* |  | |
| *助教* |  | | *电子信箱* |  | |
| 教  材 | **指定教材： 《数据挖掘方法与应用》（第1版），徐雪琪编著，清华大学出版社，2020。**  *参考书目：*《数据挖掘概念与技术》(第3版)，[(加)Jiawei Han；Micheline Kamber](http://book.douban.com/search/%28%E5%8A%A0%29Jiawei%20Han%3BMicheline%20Kamber)，机械工业出版社，2012。  《SPSS Modeler数据挖掘方法与应用》(第3版)，薛薇编著，电子工业出版社，2020。 | | | | | |
| 考试安排 | 总成绩构成比例：平时 30 ％ ，期末考试 70 %。 | | | | | |
| 教学目的 | 本课程坚持以应用为导向、以方法为基础，通过循序渐进地讲解数据挖掘可使用的工具、数据存储及分析环境、原始数据可能存在的问题及相应的预处理方法、数据挖掘经典算法等相关知识，使学生对数据挖掘有整体的认识和了解。此外，本课程以解决问题为目的，结合实例阐述使用IBM SPSS Modeler和R软件进行数据挖掘的方法与步骤，便于学生更好地理解和掌握，并能至少熟练使用一种数据挖掘分析工具分析解决实际问题。通过本课程的学习，希望能提升学生分析问题、解决问题的能力。 | | | | | |
| 课程要求 | 上课不迟到、不早退，认真听讲；  课堂讨论准备充分、发言积极；  课后思考分析题与上机练习题认真完成，尽量做到课前预习和课后及时复习；  期末提交一份数据挖掘应用分析相关的大作业，并附上分析数据及相应数据流或R代码。  以上都作为平时成绩的依据。 | | | | | |
| 课程思政 | 本课程教学过程中，力求将思政教育内容以润物细无声的方式渗透到专业授课过程中，以期实现知识传授与价值引领的有机统一。具体目标如下：  一、培养学生的爱党、爱国情怀和家国情怀；  二、培养学生的宪法法治意识；  三、培养学生奋勇攀登的学习精神和为人民服务意识；  四、培养学生的责任意识和职业素养；  五、培养学生的创新理念和创新精神。 | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **周数** | **时数** | **教学主题** | **教学内容（含课程思政内容）** | **教学方式[[3]](#footnote-3)** | **预习要求** | **作业要求** |
| 1 | 3 | 数据挖掘概述 | **第1章 数据挖掘概述**  一、课程介绍  二、参考教材、阅读书目说明  三、大数据概念、特点、应用现状及发展前景。  课程思政：分析我国大数据及其应用的发展现状，培养学生的爱党、爱国情怀和家国情怀。  1.1数据挖掘的产生与发展  1.2数据挖掘过程 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 2 | 3 | 数据挖掘概述  数据挖掘工具 | 1.3数据挖掘功能与使用技术  课程思政：结合大数据时代隐私保护的问题培养学生的宪法法治意识；结合大数据时代特征，培养学生的创新理念和创新精神。  1.4 数据挖掘应用  **第2章数据挖掘工具**  2.1 Weka | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 3 | 3 | 数据挖掘工具  数据与数据平台 | 2.2 IBM SPSS Modeler  2.3 R语言  2.4 Python语言  课程思政：分析我们可以做什么，我们需要准备什么，以已经毕业的学长为例，培养学生奋勇攀登的学习精神和为人民服务意识。  **第3章 数据与数据平台**  3.1 数据类型  3.2 关系型数据库 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 4 | 3 | 数据与数据平台  数据预处理 | 3.3 NoSQL数据库  3.4数据仓库与大数据平台  （大数据平台可视教学时间与学生基础，选讲）  **第4章 数据预处理**  4.1数据预处理概述 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 5 | 3 | 数据预处理 | 4.2数据清洗  4.3数据集成  4.4数据变换  4.5数据归约  课程思政：唯有高质量的数据，才有可能挖掘到有用的知识，结合数据预处理的特点，培养学生的责任意识和职业素养。 | 线下教学 | 阅读教材相关内容 | 练习多维数据模型的构建及OLAP操作 |
| 6 | 3 | 关联分析 | **第5章关联分析**  5.1 关联分析概述  5.2 Apriori算法  5.3 强关联规则的悖论  5.4 基于IBM SPSS Modeler的应用 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 7 | 3 | 关联分析  决策树 | 5.5 基于R语言的应用  **第6章 决策树**  6.1决策树概述  6.2 ID3算法 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 8 | 3 | 决策树 | 6.3 C5.0算法  6.4基于IBM SPSS Modeler的应用  6.5基于R语言的应用 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 9 | 3 | 贝叶斯分类 | **第7章 贝叶斯分类**  7.1 贝叶斯分类概述  7.2 朴素贝叶斯分类  7.3 TAN贝叶斯分类  7.4基于IBM SPSS Modeler的应用 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 10 | 3 | 贝叶斯分类  神经网络 | 7.5基于R语言的应用  **第8章 神经网络**  8.1 神经网络概述  8.2 BP神经网络 | 线下教学 | 阅读教材相关内容 | 教材上配套的有关题目 |
| 11 | 3 | 神经网络 | 8.3 卷积神经网络（可视教学时间，选讲）  8.4基于IBM SPSS Modeler的应用  8.5基于R语言的应用 | 线下教学 | 阅读教材相关内容 | 准备下次实验内容 |
| 12 | 3 | 数据预处理相关实验  关联分析相关实验 | 基于IBM SPSS Modeler的数据预处理  基于IBM SPSS Modeler进行事实表购物数据集关联规则挖掘练习 | 线下教学 | 阅读教材相关内容 | 练习本次实验内容 |
| 13 | 3 | 关联分析相关实验 | 基于IBM SPSS Modeler的事务表观影数据集关联挖掘  基于R语言的关联挖掘 | 线下教学 | 阅读教材相关内容 | 练习本次实验内容 |
| 14 | 3 | 决策树相关实验 | 基于IBM SPSS Modeler的决策树分类  基于R语言的决策树分类 | 线下教学 | 阅读教材相关内容 | 练习本次实验内容 |
| 15 | 3 | 贝叶斯分类相关实验 | 基于IBM SPSS Modeler的贝叶斯分类  基于R语言的贝叶斯分类 | 线下教学 | 阅读教材相关内容 | 练习本次实验内容 |
| 16 | 3 | 神经网络相关实验 | 基于IBM SPSS Modeler的神经网络  基于R语言的神经网络 | 线下教学 | 阅读教材相关内容 | 考试复习 |

1. 课程类型： 普通共同课、学科共同课、专业核心课、专业选修课、通识选修课 [↑](#footnote-ref-1)
2. 课程性质：理论课、独立设置的实验课、含课内实验的理论课、实践课 [↑](#footnote-ref-2)
3. 教学方式： 线下教学、线上自主学习、社会实践（外出考察）、其他 [↑](#footnote-ref-3)