# 《数据挖掘与应用》教学大纲

## 一、课程基本信息

1. 课程名称：数据挖掘与应用

2. 课程体系/类别：专业拓展课

3. 学分学时：3学分，58学时（含课堂教学28学时，实践教学30学时）

4. 适用专业：人工智能技术应用

5. 考核方式：考查

6. 先修课程：Python程序设计、机器学习

7. 后续课程：无

8. 开课单位：人工智能学院

## 二、课程简介

本课程系统讲述数据挖掘工具的实际应用，教授数据分析与处理的实用技能，以及挖掘算法的构建与实践。通过理论与实践相结合的教学方式，学生将掌握数据挖掘的核心概念、算法和工具使用，理解数据预处理、特征提取、模型开发等关键技术。课程强调实践操作，通过项目驱动的教学方法，培养学生的动手能力和解决复杂问题的能力，为学生未来在数据挖掘领域的职业发展打下坚实基础。

## 三、课程思政

本课旨在培养学生的社会责任感和职业道德，强调在数据挖掘实践中遵守法律法规，尊重数据隐私，注重数据安全，防止数据泄露和滥用。同时，课程鼓励学生保持科学精神和创新思维，通过团队合作和项目实践，提升沟通能力和解决复杂问题的能力。此外，课程还注重培养学生的终身学习意识，激发他们为社会服务和贡献国家的热情，同时教育学生在数据分析中保持客观和批判性思维，以及在算法设计中追求公平正义，减少偏见，从而在数据挖掘领域成为具有高度社会责任感和专业素养的人才。

## 四、课程目标

通过本课程学习，学生应达到的预期教学目标如下：

课程目标1：了解数据挖掘的基本概念、发展历程、应用领域以及Web挖掘技术的一般流程和数据挖掘常用工具；

课程目标2：理解数据挖掘常用算法的核心思想、各算法的联系与区别以及适用场景，并能够根据具体的业务场景选择构建合适的数据挖掘模型；

课程目标3：掌握数据挖掘常用分类、聚类、回归和关联分析等算法的模型构建与训练流程、性能评估、超参数调优以及推理与部署等工程实践能力

课程目标4：培养学生的创新创造与知识迁移能力，引导学生关注数据挖掘领域的最新研究进展和技术趋势，通过阅读行业报告和跟进行业发展趋势，培养学生的自主学习能力和终身学习意识。

## 五、教学内容与课程目标的关系

**（1）理论教学**

本门课的主要教学内容与要求，以及学时分配如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **教学内容** | **教学要求** | **学时** | **教学方式** | **课程思政**  **内容** | **对应**  **课程**  **目标** |
| 1 | 1、数据挖掘基础  （1）数据预处理  （2）数据挖掘工具  （3）Web挖掘技术 | （1）了解数据挖掘的基本概念、产生与发展、知识表示、知识发现以及数据挖掘的应用  （2）理解Web内容挖掘方法与Web访问信息方法  （3）掌握RapidMiner、Sklearn和PySpark等常用数据挖掘工具的使用 | 4 | 讲授 | 培养学生的数据预处理与分析能力，通过Web挖掘技术理解网络信息价值，强化网络素养与数据安全意识 | 目标1  目标4 |
| 2 | 2、关联分析  （1）关联规则的基本概念  （2）关联规则的挖掘过程  （3）关联规则的Apriori算法  （4）关联规则的FP-Growth算法 | （1）了解关联规则的基本概念与挖掘过程  （2）理解Apriori和FP-GRowth等经典关联规则算法的基本原理和实现流程  （3）掌握使用基于Apriori和FP-GRowth关联规则算法的工具实现关联分析等任务实践技能 | 8 | 讲授 | 培养学生的关联分析技能，掌握数据内在联系，提升逻辑推理与问题解决能力，强化数据驱动决策的科学精神 | 目标2  目标3 |
| 3 | 3、回归算法  （1）线性回归  （2）逻辑回归 | （1）了解回归算法的基本概念和适用场景  （2）理解线性回归和逻辑回归等算法的核心思想和实现原理  （3）掌握线性回归和逻辑回归等算法的技能实践 | 6 | 讲授 | 培养学生的回归分析能力，强化数据分析与科学决策的实践技能，提升服务社会和经济建设的责任感 | 目标2  目标3 |
| 4 | 4、分类算法  （1）分类基本概述  （2）K-近邻分类算法  （3）贝叶斯分类  （4）决策树分类  （5）SVM支持向量机  （6）神经网络分类算法  （7）集成学习 | （1）了解分类算法的基本概念和适用场景  （2）理解K-近邻、贝叶斯、决策树与集成学习、支持向量机、神经网络等算法的基本原理和实现方式  （3）掌握使用分类算法实现分类预测任务的实践技能 | 6 | 讲授 | 培养学生的分类算法应用能力，强化数据驱动的决策思维，提升解决复杂问题的能力，增强创新实践和服务社会的意识 | 目标2  目标3 |
| 5 | 5、聚类方法  （1）聚类基本概念  （2）划分聚类方法  （3）层次聚类方法  （4）密度聚类方法 | （1）理解聚类算法的基本概念和适用场景  （2）理解划分聚类、层次聚类和密度聚类等典型聚类算法的基本思想和实现流程  （3）掌握常用聚类算法的任务实践 | 4 | 讲授 | 培养学生在多样化数据中发现规律的能力，激发他们在复杂问题中寻找创新解决方案的动力 | 目标2  目标3 |

表1：教学内容与目标的关系

**（2）实践教学**

其中，实践环节主要包含以下项目：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **实践名称** | **实践内容** | **学时** | **教学方式** | **课程思政**  **内容** |
| 1 | 基于互联网在线用户行为数据预处理技术V1.0 | （1）数据导入操作  （2）数据转换处理  （3）数据归一化处理  （4）替换缺失值处理  （5）数据文件导出 | 2 | 线上实验 | 培养数据意识与责任感，提升信息质量与分析准确性，强化在数据驱动决策中的伦理与隐私保护意识 |
| 2 | 基于Python的Sklearn数据挖掘项目实战 | （1）数据获取  （2）创建数据库  （3）数据预处理  （4）模型构建  （5）模型评估与选择  （6）保存模型  （7）任务结果显示 | 2 | 线上实验 | 培养数据思维与实践能力，提升数据分析与问题解决能力，强化科技在服务社会决策中的责任与担当 |
| 3 | 中文文本挖掘实战 | （1）数据库导入  （2）jieba分词  （3）词汇处理  （4）TF-IDF算法抽取特征向量  （5）结果展示 | 2 | 线上实验 | 培养语言处理与文化自信，提升对中文信息的分析与理解能力，强化科技在传承和弘扬中华文化中的应用价值 |
| 4 | 基于协同过滤算法的电影推荐 | （1）数据加载  （2）基于用户推荐算法的电影推荐模型构建  （3）基于用户的推荐列表生成  （4）推荐模型评估 | 2 | 线上实验 | 培养创新思维与服务意识，提升数据分析与用户需求理解能力，强化科技在提升用户体验与文化产品推广中的作用 |
| 5 | 关联规则分析 | （1）原始数据导入  （2）原始数据预处理  （3）关联规则函数定义  （4）关联规则数目分析图生成 | 2 | 线上实验 | 培养逻辑分析与决策能力，增强从大量数据中发现有价值信息的能力，强化在商业智能与决策支持中的应用意识 |
| 6 | 基于某城市社区活动关联规则分析 | （1）数据加载  （2）特征预处理  （3）关联规则分析  （4）文件保存 | 2 | 线上实验 | 培养社会服务意识与数据分析能力，提升社区服务优化与居民需求响应能力，强化科技在促进社区和谐发展中的应用价值 |
| 7 | 基于逻辑回归技术预测心脏病二次发作率V2.0 | （1）原始数据可视化  （2）特征预处理  （3）逻辑回归模型训练  （4）模型预测与评分  （5）交叉验证调参  （6）心脏病发作预测 | 2 | 线上实验 | 培养健康关怀与数据分析技能，提升对健康问题的关注与预防意识，强化科技在公共卫生领域中的应用责任 |
| 8 | 基于线性回归技术预测某社区新客户的热燃油用量V2.0 | （1）原始数据可视化  （2）特征预处理  （3）线性回归模型训练  （4）线性回归模型评估  （5）热燃油用量预测与结果可视化 | 2 | 线上实验 | 培养能源管理意识与预测分析能力，提升节能减排与资源优化配置的实践技能，强化科技在支持可持续发展中的作用 |
| 9 | 基于ID3算法的决策树分类实战 | （1）数据模拟生成  （2）决策树模型构建  （3）模型预测结果输出 | 2 | 线上实验 | 培养科学精神和严谨态度，强化数据驱动决策的意识，提升解决实际问题的能力 |
| 10 | 基于分类技术对某零售商产品精准广告投放营销V3.0 | （1）数据加载  （2）特征预处理  （3）决策树分类模型构建  （4）决策树分类模型评估  （5）决策树结构图构建  （6）决策树分类模型应用 | 2 | 线上实验 | 培养数据素养和创新思维，强化数据分析与应用能力，促进营销策略的科学化和精准化 |
| 11 | 基于文本挖掘技术对垃圾短信、邮件精准检测V1.0 | （1）使用TF-IDF算法抽取短信特征向量  （2）逻辑回归分类模型训练与预测  （3）逻辑回归分类模型交叉验证与评估 | 2 | 线上实验 | 培养社会责任感和法律意识，提升信息安全意识，维护网络环境的清洁与秩序 |
| 12 | 基于聚类分析方法的患冠心病率状况探索V1.0 | （1）K-means聚类模型构建  （2）聚类模型预测结果可视化  （3）聚类模型效果评估 | 2 | 线上实验 | 培养健康关怀与数据分析技能，提升对健康问题的关注与预防意识，强化科技在公共卫生领域中的应用责任 |
| 13 | 神经网络实战 | （1）数据加载与数据集划分  （2）数据标准化处理  （3）神经网络分类模型构建  （4）神经网络模型训练  （5）分类结果预测  （6）模型效果可视化评估 | 2 | 线上实验 | 培养创新意识和实践能力，强化人工智能技术在解决复杂问题中的应用，提升科技服务社会的能力 |
| 14 | 基于分类技术对某零售商产品精准广告投放营销V4.0 | （1）数据加载  （2）特征预处理  （3）神经网络分类模型构建  （4）神经网络分类模型训练  （5）模型预测与效果评估 | 2 | 线上实验 | 培养数据素养和创新思维，强化数据分析与应用能力，促进营销策略的科学化和精准化 |
| 15 | 基于神经网络算法的职业足球冠军赛的球员人选分析 | （1）数据加载与特征预处理  （2）神经网络分类模型训练与预测  （3）神经网络分类模型应用 | 2 | 线上实验 | 培养数据驱动的决策思维和模型应用能力，激发体育数据分析兴趣，促进科技与体育的深度融合，助力体育事业创新发展 |

表2：实践内容与学时分配

## 六、课程教学方法

本课程侧重学生个人学习能力的培养，老师主讲理论基础知识，实验环节演变为技术的指导者，帮助学生解决实践操作中的问题，学生自主上机操作完成实操案例。学习模式采取“教师主讲理论知识，人工智能实验平台支撑，助教辅助，梯形团队管理”，主讲老师大班理论、重难点讲解，助教小班作业辅导。

1. 讲师课堂讲授

(1) 采用启发式教学，激发学生主动学习的兴趣，培养学生对专业的认识，引导学生明确学习目标。

(2) 采用案例式教学。理论教学与工程实践相结合，增强学生对专业的学习兴趣。

(3) 采用互动式教学。课内讨论和课外答疑相结合。

2. 助教辅导答疑

(1) 跟班辅导。做好考勤工作，随时关注学生学习状态，维持课堂记录，了解和记录学生知识掌握情况。

(2) 独自辅导。了解和记录学生任务和作业完成情况，主动巡查并进行答疑。

(3) 协助批改作业、试卷。

## 七、课程考核要求

1. 课程目标达成度

课程目标达成度评价包括课程分目标达成度评价，以及课程总目标达成度评价，具体计算方法如下：





根据本课程的教学内容、知识掌握的要求、以及前面几年的课程考核等综合因素考虑，本课程总目标达成度设置为0.6。

2. 课程考核细则

检验课程目标达成度，评价学生学习成果达成度。本课程成绩以百分计，满分100分。考核环节包含平时成绩、期末考核（平台）和期末考核（学校）等。各考核环节所占分值比例可根据具体情况微调，建议值及考核细则如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **考核**  **环节** | | **成绩** | **考核/评价范围与标准** | **考核方式** | **映射课程目标** |
| 平时成绩 | 出勤 | 10 | 根据学生出勤情况打分，三次缺勤，该出勤考核为0分；五次缺勤，平时成绩为0分。 | 过程评分 | 目标1 |
| 平台 | 10 | 根据平台课程学习情况（如进度、时长、正确率）进行考核，如果进度低于80%，则此项考核成绩为0分。 | 过程评分 | 目标2  目标3 |
| 作业 | 20 | （1）教学过程中包含课堂讨论及小组学习等环节，课堂表现考核学生的课堂参与程度；  （2）作业项考核学生的作业是否及时、独立地完成。 | 过程评分 | 目标2  目标4 |
| 期末考核  （平台） | | 20 | 考核范围为课程大纲考核范围内的所有内容。  注：允许平台考核不高于5%的超纲范围 | 平台考核 | 目标3  目标4 |
| 期末考核  （学校） | | 40 | （1）采用笔试，按照学校教学计划进行，考核范围为课程大纲包含的全部内容知识点和能力要求。  （2）期末考试具有一票否决权，如果期末笔试的卷面成绩在50分以下，则该课程直接不及格，课程成绩为期末笔试卷面成绩；如果期末笔试卷面成绩等于或高于50分，按照分值比例归于课程成绩。 | 笔试 | 目标1  目标2  目标4 |

表3：课程考核环节及评分细则

## 八、教材与参考书

教材：

[1] 《数据挖掘：概念与技术》.[美] [Jiawei](http://search.dangdang.com/?key2=Jiawei&medium=01&category_path=01.00.00.00.00.00" \t "https://product.dangdang.com/_blank)著. 机械工业出版社.2022.06

[2] 《数据挖掘实用案例分析（第2版）》[中] 赵卫东、董亮 等著. 清华大学出版社.2022.03

参考资料：

[1] 人工智能实验平台中的《数据挖掘与应用》课程