

《数据挖掘》课程教学大纲

课程英文名	Data Mining				
课程代码	B050737s	课程类别	专业课	课程性质	专业选修
学 分	3		总学时数	48	
开课学院	计算机学院		开课基层教学组织	认知课程组	
面向专业	大数据与智能决策（管理科学与工程类）、计算机科学英才班（计算机科学与技术）、计算机科学与技术、计算机科学与技术（第二学士学位专业）、软件工程、智能财务（会计学）、智能财务（软件工程）、智能计算与数据科学（计算机科学与技术）		开课学期	第 2、4、5 学期	

一、 课程目标

本课程属计算机科学与技术及相关专业的选修课。通过本课程的学习，让学生在学习期间掌握数据挖掘理论以及如何用数据挖掘来解决实际问题，了解某个数据挖掘解决方案对特定问题是否切实可行，学习知识发现的流程，利用基本的统计技术评估数据挖掘的结果。结合新时代中国特色社会主义建设的背景，深入了解我国在数据挖掘领域的优势和不足，增强学生爱国情怀，激发学生历史使命感，树立利用数据挖掘技术造福社会的志愿，投身于中国梦的伟大工程。重点掌握几种数据挖掘策略及每种策略的适用时机，通过该课程的学习能独立采用适当数据挖掘的技术来建立模型进而解决实际问题，为学生今后从事相关领域的工作打下较坚实的基础。

通过本课程各项教学活动的实施，达到以下课程目标：

课程目标 1：掌握数据挖掘的流程、基本原理与技术（算法），掌握我国在此领域面的前沿进展和需求，树立强烈的爱国主义使命感与责任心；

课程目标 2：数据挖掘领域涉及很多现实生活和生产过程中的实际问题，能够将所学的数据挖掘模型及相关算法应用于实际问题求解。

二、 课程目标与毕业要求对应关系

本课程的课程目标与计算机科学与技术的毕业要求及其指标点的对应关系如表 1 所示。

表 1. 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
------	-----	-----------

毕业要求 1：工程知识：能够掌握数学、自然科学、工程基础、计算机软件系统和计算机硬件体系知识，并应用在计算机相关领域的复杂工程问题的解决方案中。	1-2 能够对模型进行分析，并利用模型解决问题。	目标 1：1.0
毕业要求 5：使用现代工具：能够针对复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对复杂工程问题的预测与模拟，并能够理解其局限性。	5-2 能够开发、选择和使用恰当的技术和资源对计算机相关领域复杂工程问题进行预测与模拟。	目标 2：1.0

本课程的课程目标与软件工程专业的毕业要求及其指标点的对应关系如表 2 所示。

表 2. 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
毕业要求 1：工程知识：能够掌握数学、自然科学、工程基础和计算机软件系统体系知识，并应用在软件工程相关领域的复杂工程问题的解决方案中。	1-2 能够应用工程基础及软件工程专业知识对模型进行分析，并利用模型解决问题。	目标 1：1.0
毕业要求 5：使用现代工具：能够针对软件工程领域复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对复杂工程问题的预测与模拟，并能够理解其局限性。	5-2 能够开发、选择和使用恰当的技术和资源对软件工程相关领域复杂工程问题进行预测与模拟。	目标 2：1.0

本课程的课程目标与智能财务专业的毕业要求及其指标点的对应关系如表 3 所示。

表 3. 课程目标与毕业要求对应关系

毕业要求	课程目标及支撑权重
毕业要求 1：工程知识:掌握数学、自然科学、工程基础、财务领域基础知识和软件工程专业领域知识，并能应用在智能财务软件领域复杂工程问题的解决方案中。	目标 1：1.0
毕业要求 5：使用现代工具:能够针对智能财务软件领域复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对复杂工程问题的预测与模拟，并能够理解其局限性。	目标 2：1.0

本课程的课程目标与智能计算与数据科学专业的毕业要求及其指标点的对应关系如表 4 所示。

表 4. 课程目标与毕业要求对应关系

毕业要求	课程目标及支撑权重
毕业要求 1：工程知识:能够掌握数学、自然科学、工程基础、计算机软硬件知识、人工智能、智能计算和数据科学的基础理论及专业知识，并应用在人工智能、智能计算和大数据专业领域及其他相关交叉领域的复杂工程问题的解决方案中。	目标 1：1.0
毕业要求 5：使用现代工具:能够针对复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对复杂工程问题的预测与模拟，并能够理解其局限性。	目标 2：1.0

本课程的课程目标与大数据与智能决策专业的毕业要求及其指标点的对应关系如表 5 所示。

表 5. 课程目标与毕业要求对应关系

毕业要求	课程目标及支撑权重
毕业要求 2：问题分析:具有较为全面的系统观念，能够应用数学、管理科学和工程科学的基本原理，识别、表达、并通过文献研究分析数	目标 1：1.0

字化工程问题，获得有效结论。	
毕业要求 3：设计/开发解决方案:参与数字化工程项目实践，运用工业工程先进的理论 和方法，初步具备数字化工程的规划和设计、运营和管理等能力，并能够在设计环节中体现 创新意识，考虑社会、健康、安全、法律、文化以及环境等因素。	目标 2：1.0
毕业要求 5：使用现代工具:能够针对工业工程尤其是数字化工程问题，开发、选择与使用大数据、人工智能、各种算法与计算工具，并理解其局限性。	目标 2：1.0

三、 课程目标与教学内容和方法的对应关系

《数据挖掘》教学内容对课程目标的支撑关系、教学方法如表 6 所示：

表 6. 课程目标与教学内容、教学方法的对应关系

教学内容	教学方法	课程目标
1. 数据挖掘概述	课堂讲授、课后自学、文献查阅	1
2. 数据预处理	课堂讲授、提问、课内上机	1, 2
3. 分类和预测	课堂讲授、课堂讨论、课内上机、文献查阅	1, 2
4. 关联分析	课堂讲授、课堂讨论、课内上机、文献查阅	1, 2
5. 聚类分析	课堂讲授、课堂讨论、课内上机、文献查阅	1, 2
6. 异常检测	课堂讲授、课堂讨论、课内上机、文献查阅	1, 2

课程教学的详细内容与要求如下：

1. 概论

(1) 教学内容：

- 数据挖掘基本概念；数据挖掘处理的数据对象；数据挖掘主要功能；
- 数据挖掘的主要问题；
- 数据挖掘过程与功能，数据挖掘的应用和面临的问题。

(2) **教学重点：**数据挖掘的起源，数据挖掘过程与功能，以及面临的主要问题。

(3) **教学难点：**数据挖掘面临的问题，解决问题思路。

(4) **教学要求：**了解数据挖掘的过程和功能，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。

思政融合点 1：由数据挖掘的重要意义，引导学生了解我国在基于数据挖掘改善人民生活方面的应用和成就。通过中美之间在科技领域的竞争，激发学生的爱国主义热情、自豪感与使命感。

2. 数据预处理

(1) 教学内容：

- 数据及数据的描述，数据清理、数据集成、数据变换和数据归约。
- 数据预处理的目的是意义；
- 不同数据源的数据进行合并；
- 对数据进行变换；目前数据预处理发展及研究动态。

(2) **教学重点：**数据的各种预处理方法

(3) **教学难点：**数据预处理的重要性和必要性。

(4) 教学要求：能够对特定的数据采用合理的预处理方法。

思政融合点 2：2015 年大数据正式成为国家战略。介绍国务院的《促进大数据发展行动纲要》，工信部的《大数据产业发展规划（2016—2020 年）》，激发学生了解我国对大数据产业的总体设计和统筹，并学习我国大数据产业发展现状及面临的形势，引导学生思考数据的复杂性、可用性，以及可能引发的社会问题（比如隐私和信息泄露等）。

3. 分类和预测

(1) 教学内容：

- 类和预测的基本概念；决策树归纳分类；
- 贝叶斯分类；神经网络分类；
- 基于规则的分类；惰性学习方法；
- 分类器的准确性的度量与估计。

(2) 教学重点：常用的分类及预测算法（或模型）。

(3) 教学难点：分类器的准确性的度量与估计。

(4) 教学要求：掌握常用的分类及预测算法（或模型），掌握分类和预测分析如何通过数据挖掘软件实现。

思政融合点 3：引导学生利用数据挖掘的相关技术，设计可以改善人民生活的预测及推荐系统，培养学生的创新精神和科学精神，引导学生积极思考如何让知识和技术发展有益于人民、有益于国家。

4. 关联分析

(1) 教学内容：

- 频繁项集、关联规则；
- 可伸缩的频繁项集挖掘方法；
- 各种类型的关联规则和基于约束的关联挖掘。

(2) 教学重点：关联规则的基本思想；关联规则挖掘的应用背景；掌握常用的频繁项集挖掘算法，包括 Apriori、FP 树的频繁项集方法；多层关联规则、多维关联规则的基本思想；

(3) 教学难点：关联规则的基本思想，Apriori、FP 树的频繁项集方法。

(4) 教学要求：要求了解关联规则的基本思想、概念和意义以及关联规则挖掘的应用背景；掌握常用的频繁项集挖掘算法。

思政融合点 4：引导学生利用关联分析技术，对 COVID-19 和国家经济行业走向、人才走向的关联性进行分析，引导学生积极思考如何择业让国家有能力应对国际突发事件。

5. 聚类分析

(1) 教学内容：

- 聚类分析的基本概念
- 聚类分析中的基本数据类型
- 主要聚类方法：K-Means、k 中心点方法、BIRCH 等。

- (2) **教学重点**：常用的聚类算法，包括 K-Means、k 中心点方法、BIRCH 等。
- (3) **教学难点**：相似度（或相异度）的度量方法和聚类算法。
- (4) **教学要求**：掌握常用的聚类算法，包括 K-Means、k 中心点方法、BIRCH 等；掌握聚类分析如何通过数据挖掘软件实现；了解聚类分析的研究动态。

思政融合点 5：引导学生运用聚类技术，能够在民生相关的政治、文化、金融、教育等领域数据中发现具有共性的小团体，从而促进新的综合体的成长和壮大，造福人民。

6. 异常检测

(1) 教学内容：

- 异常检测的基本概念
- 统计方法；检测方法特性；检测方法评估
- 主要检测方法：基于邻近度的方法、基于聚类的方法、基于重构的方法等。

(2) **教学重点**：常用的异常检测方法，包括基于邻近度的方法、基于聚类的方法、基于重构的方法等。

(3) **教学难点**：各种异常检测方法的灵活运用。

(4) **教学要求**：掌握常用的异常检测算法，包括基于邻近度的方法、基于聚类的方法、基于重构的方法等；掌握检测方法特性和异常检测评估。

思政融合点 6：引导学生运用异常检测技术，能够在公共卫生和安防领域的舆情数据中及时发现并报告异常事件，从而为国家的公共安全做出贡献。

四、 实践环节及要求

1. 实践项目和基本要求

课程包含 12 学时的课内上机，以及 12 学时的课外上机。在课堂中讲授上机任务的要求为主，学生在课内(课外)上机时完成特定的数据挖掘问题。对具体的数据，完成以下的操作：数据预处理，进行探索性分析，计算数据的统计汇总并可视化，对数据进行频繁相集挖掘，对数据分类，对数据聚类。通过实验项目，巩固所学知识，使得学生能够开发、选择和使用恰当的数据挖掘技术和资源对相关领域复杂数据工程问题进行处理和预测，培养学生独立分析问题、解决问题的能力。

表 7. 课内上机安排

序号	实验项目	时数	每组人数	内容提要	实验要求
1	数据预处理和汇总计算	3学时	1~5	对给出的真实数据进行合适的 数据预处理，并对数据进行统计 汇总计算和可视化。	编程实现并可视化
2	频繁相集挖掘	3学时	1~5	实现频繁相集的基本算法并用于 实际数据	编程实现并分析结果
3	数据的分类	3学时	1~5	对于给定数据用分类方法分析 数据，并给出分析及结果。	编程实现并分析结果

4	数据的聚类分析	3学时	1~5	对于给定数据用聚类方法分析数据，并给出分析及结果。	编程实现并分析结果
---	---------	-----	-----	---------------------------	-----------

2. 实验报告基本要求

实验报告至少包含以下几个部分：(1)实验目的；(2)实验环境；(3)实验内容；(4)实验结果及结果分析；(5)实验总结。

五、 与其它课程的联系

先修课程：数据结构、程序设计基础、数据库原理及应用（建议）

后续课程：无

六、 学时分配

总学时为 48 学时，讲课 36 学时，课内上机 12 学时，课外上机 12 学时和自学 12 学时，如表 8 所示。

表 8. 学时分配表

教学内容	讲课时数	实验时数	实践学时	课内上机时数	课外上机时数	自学时数	习题课	讨论时数
1. 数据挖掘概述	3					2		
2. 数据预处理	6			3	3	2		
3. 分类和预测	9			3	3	2		
4. 关联分析	6			3	3	2		
5. 聚类分析	9			3	3	2		
6. 异常检测	3					2		
合计	36			12	12	12		
总计	计划 48 学时（讲课 36 学时+课内上机 12 学时+课外上机 12 学时）+自学 12 学时							

七、 课程目标达成途径及学生成绩评定方法

1. 课程目标达成途径

各个课程目标的达成途径如表 9 所示，但不仅限于此。

表 9. 课程目标与达成途径

课程目标	达成途径
课程目标 1： 掌握数据挖掘的流程、基本原理与技术（算法），掌握我国在此领域面的前沿进展和需求，树立强烈的爱国主义使命感与责任心。	以引导式、启发式和总结式教学方法为主，通过重点/难点内容讲解、进行随堂提问、算法程序讲解、课后自学、文献查阅、课堂讨论、分析对比、总结报告等各种方式等模式，帮助学生学习相关知识，建立树立强烈的爱国主义使命感与责任心。
课程目标 2： 数据挖掘领域涉及很多现实生活和生产过程中的实际问题，能够将所学的数据挖掘模型及相关算法应用于实际问题求解。	通过课内上机的任务体现：给定真实数据，提出数据挖掘的具体要求和注意事项，学生按要求编程完成具体数据挖掘任务。帮助学生运用所学知识处理实际问题。

2. 学生成绩评定方法

本课程为考查课程。课程采用形成性评价与终结性评价相结合的评价方法，学期总评成绩由两

部分构成：平时成绩，占比 50%；期末成绩，占比 50%。各部分的具体评价环节、关联课程目标、评价依据及方法和在总成绩中的占比，如表 10 所示。

表 10. 课程考核与成绩评定方法

考核项目	考核内容	考核关联的课程目标	占考核项目成绩比例	占总评成绩的比重
平时成绩	课程思政实践	1	20%	50%
	上机报告	1	40%	
	项目报告	1, 2	40%	
期末成绩	期末大作业	1, 2	100%	50%
总评成绩		1, 2		100%

各考核内容的详细评分标准见表 11 所示。

表 11. 考核内容详细评分标准

考核内容		评分标准			
		90-100	75-89	60-74	<60
课程思政实践		报告条理清晰，文字流畅，字数≥4000，参考文献数量≥8 且相关性强；内容完整且材料丰富，体现强烈的使命感、责任心与民族自豪感	报告条理清楚，字数≥3000，参考文献数量≥5 且相关性较好；内容完整，材料不够丰富，能体现学生的使命感、责任心与民族自豪感	报告有一定条理，字数≥1000，参考文献数量≥2 且基本相关；内容基本完整但材料较少，能体现学生的使命感与民族自豪感	报告字数<1000，参考文献数量<2；内容少，或有抄袭现象，体现不出学生的使命感与民族自豪感
上机报告	基础知识	根据学生答题正确率进行评判			
期末大作业	代码实现	程序运行流畅，功能完善，性能好	程序运行正常，功能基本实现，但不够完善、优化等	程序运行基本正常，有少量 bug，功能实现至少达到 80%	程序运行 bug 多，功能实现低于要求的 80%
	文档质量	格式规范，图表清晰美观，内容完整具体有条理，字数≥4000，不抄袭	格式较规范，图表清晰，内容较完整、有条理，字数≥3000，不抄袭	格式基本规范，内容基本完整，字数≥1000，不抄袭	格式不规范，内容不完整，字数<1000，或有抄袭现象
	原创性	代码实现为原创，报告无抄袭	代码实现大部分为原创，报告无抄袭	代码实现 60% 以上为原创，报告无抄袭	代码实现原创部分少于 60%，报告存在抄袭
项目报告	主题选择	所选主题紧密围绕“中国正能量”	所选主题较好地围绕“中国正能量”	所选主题一定程度地围绕“中国正能量”	所选主题未能围绕“中国正能量”
	代码实现	程序运行流畅，功能完善，性能好	程序运行正常，功能基本实现，但不够完善、优化等	程序运行基本正常，有少量 bug，功能实现至少达到 80%	程序运行 bug 多，功能实现低于要求的 80%
	界面	界面布局美观，	界面布局合理，基本符合	界面布局较为合理，部	界面布局设计不合理

	设计	符合应用设计规范	应用设计规范	分模块不符合设计规范	
	报告质量	格式规范，图表清晰美观，内容完整具体有条理；对设计思想、实现方法、运行结果有较完整、准确的描述	格式较规范，图表清晰，内容较完整、有条理；对设计思想、实现方法、运行结果有较完整的描述	格式基本规范，内容基本完整；对设计思想、实现方法、运行结果有描述	格式不规范，内容不完整
	原创性	代码实现、界面设计均为原创，报告无抄袭	代码实现、界面设计大部分为原创，报告无抄袭	代码实现、界面设计60%以上为原创，报告无抄袭	代码实现、界面设计原创部分少于60%，报告存在抄袭

八、教学资源

表 12. 课程的基本教学资源

资源类型	资源
教材	Pang-Ning Tan, Michael Steinbach著,《数据挖掘导论》(英文版.原书第2版),机械工业出版社,2020 (ISBN 9787111637882)
参考书籍或文献	<p>[2] 坎塔尔季奇, 王晓海, 吴志刚译. 数据挖掘--概念模型方法和算法(第2版), 清华大学出版社, 2013.</p> <p>[3] Larose 著, 刘燕权等译, 数据挖掘方法与模型, 高等教育出版社, 2011</p> <p>[4] Witten I.H., Frank E.著, 董琳, 邱泉等译. 数据挖掘: 实用机器学习技术(原书第2版). 机械工业出版社, 2007</p> <p>[5] Sebastian Raschka, Vahid Mirjalili, Python Machine Learning - Second Edition, Packt Publishin, 2017</p> <p>[6] 伊恩·威腾, 数据挖掘: 实用机器学习工具与技术(英文版·第4版), 机械工业出版社, 2017</p>
教学文档	无

九、课程目标达成度的定量评价

在课程结束后,需要对每一个课程目标(含思政课程目标)进行达成度的定量评价,用以实现课程的持续改进。

课程目标达成度的定量评价算法:

- 1、使用教学活动(如课程思政报告、上机实验和报告,课程项目报告,演讲、课堂讨论、互动等等)成绩的得分率作为评价项目,来对某个课程目标进行达成度的定量评价;
- 2、为保证考核的全面性和可靠性,要求对每一个课程目标的评价项目选择超过两种;
- 3、根据施教情况,评价项目可以由教师自行扩展,权重比例可以由教师自行设计;
- 4、对某一个课程目标有支撑的各评价项目权重之和为1;
- 5、使用所有学生(含不及格)的平均成绩计算。

本课程的课程目标达成度的定量评价算法建议如表 13 所示,教师可根据授课方式及考核内容适当调整:

表 13. 课程目标达成度定量评价方法

课程目标	课程目标达成度评价方式
课程目标 1: 掌握数据挖掘的流程、基本原理与技术（算法），掌握我国在此领域面的前沿进展和需求，树立强烈的爱国主义使命感与责任心。	课程思政实践：0.3 上机报告：0.1 项目报告：0.3 期末大作业：0.3
课程目标 2: 数据挖掘领域涉及很多现实生活和生产过程中的实际问题，能够将所学的数据挖掘模型及相关算法应用于实际问题求解。	项目报告：0.3 期末大作业：0.7

十、 说明

本大纲规定了杭州电子科技大学计算机科学与技术及相关专业的《数据挖掘》的教学目标、教学内容、教学方法、教学要求以及考核评价方法等，承担该课程的教师必须遵照本大纲安排授课计划、实施教学过程，完成学生各个阶段与各方面的学习成果考核与评价；在学期末，需对课程目标进行达成度评价。

本课程大纲自 2021 年开始执行，生效之日原先版本均不再使用。

十一、 编制与审核

表9 大纲编制与审核信息

工作内容	责任部门或机构	负责人	完成时间
编制	智能信息处理课程组	谭敏	2022.02
审核	智能信息处理课程组	彭勇	2022.03
审定	计算机学院教学工作委员会		2022.05