

《机器学习与数据分析》专业课程教学大纲

一、课程基本信息

课程名称	(中文) 机器学习与数据分析				
	(英文) Machine Learning and Data Analytics				
课程代码	1060175	课程学分		2	
课程学时	32	理论学时	8	实践学时	24
开课学院	商学院	适用专业与年级		物流 大三年级	
课程类别与性质	专业选修课	考核方式		考试	
选用教材	《Python 数据挖掘与机器学习》，书号 978-7-302-57299-2，清华大学出版社			是否为马工程教材	否
先修课程	高等数学，线性代数，概率论，运筹学				
课程简介	<p>本课程系统介绍了数据分析与机器学习的基础方法与技术，涵盖了当前领域内的最新研究成果。课程内容包括 Python 数据分析与可视化基础、数据理解与预处理、回归分析、分类与预测、聚类分析，以及神经网络和机器学习的基本理论等。通过理论学习与实践操作相结合的方式，学生将逐步掌握从数据处理到模型构建与优化的全过程。</p> <p>本课程旨在培养学生在数据挖掘与机器学习方面的综合能力，特别是理论分析与实际应用的能力。完成课程后，学生将能够运用所学的机器学习技术解决实际问题，具备在数据驱动环境中进行深入分析和决策支持的能力。</p>				
选课建议与学习要求	选课学生应该具有一定的数学、线性代数、概率论以及运筹学基础，本课程适合商学院物流类专业，建议在第五至第八学期期间安排开设。				
大纲编写人			制/修订时间	2024.11	
专业负责人			审定时间	2024.11	
学院负责人			批准时间	2024.11	

二、课程目标与毕业要求

(一) 课程目标

类型	序号	内容
知识目标	1	掌握机器学习和数据分析的核心概念、基本原理及常用算法，如回归分析、分类、聚类和神经网络等。
	2	掌握 Python 编程语言及其在数据分析中的应用，使用 Numpy、Matplotlib 和 Scikit-learn 等库进行数据处理、分析和可视化。
技能目标	3	能够独立完成从数据收集、预处理、模型选择到结果分析的整个数据分析过程，利用所学知识解决实际问题。
	4	具备使用 Python 构建并优化机器学习模型的能力，能够根据实际需求选择合适的算法，并进行模型评估和调整。
素养目标 (含课程思政目标)	5	培养学生在数据分析与模型构建中的严谨性，注重数据的真实性、模型的合理性，以及结果分析的科学性，避免片面结论。
	6	增强学生在数据挖掘与机器学习应用中的社会责任感，理解并尊重数据隐私，确保技术应用符合社会利益和道德标准。

(二) 课程支撑的毕业要求

<p>LO1 品德修养：拥护中国共产党的领导，坚定理想信念，自觉涵养和积极弘扬社会主义核心价值观，增强政治认同、厚植家国情怀、遵守法律法规、传承雷锋精神，践行“感恩、回报、爱心、责任”八字校训，积极服务他人、服务社会、诚信尽责、爱岗敬业。</p> <p>①爱党爱国，坚决拥护党的领导，热爱祖国的大好河山、悠久历史、灿烂文化，自觉维护民族利益和国家尊严。</p> <p>④诚信尽责，为人诚实，信守承诺，勤奋努力，精益求精，勇于担责。</p>
<p>LO2 专业能力：具有人文科学素养，具备从事物流与供应链管理工作的理论知识和实践能力。</p> <p>①具有专业所需的人文科学素养。</p>
<p>LO6 协同创新：同群体保持良好的合作关系，做集体中的积极成员，善于自我管理和团队管理；善于从多个维度思考问题，利用自己的知识与实践来提出新设想。</p> <p>③能用创新的方法或者多种方法解决复杂问题或真实问题。</p>

LO7 信息应用:具备一定的信息素养,并能在工作中应用信息技术和工具解决问题。
②能够使用适合的工具来搜集信息,并对信息加以分析、鉴别、判断与整合。

(三) 毕业要求与课程目标的关系

毕业要求	指标点	支撑度	课程目标	对指标点的贡献度
LO1	①、④	H、H	6. 增强学生在数据挖掘与机器学习应用中的社会责任感,理解并尊重数据隐私,确保技术应用符合社会利益和道德标准。 5. 培养学生在数据分析与模型构建中的严谨性,注重数据的真实性、模型的合理性,以及结果分析的科学性,避免片面结论。	60%+40%=100%
LO2	①	M	1. 掌握机器学习和数据分析的核心概念、基本原理及常用算法,如回归分析、分类、聚类和神经网络等。 3. 能够独立完成从数据收集、预处理、模型选择到结果分析的整个数据分析过程,利用所学知识解决实际问题。	50%+50%=100%
LO6	③	H	2. 掌握 Python 编程语言及其在数据分析中的应用,使用 Numpy、Matplotlib 和 Scikit-learn 等库进行数据处理、分析和可视化。	100%
LO7	②	M	4. 具备使用 Python 构建并优化机器学习模型的能力,能够根据实际需求选择合适的算法,并进行模型评估和调整。	100%

三、实验内容与要求

(一) 各实验项目的基本信息

序号	实验项目名称	实验类型	学时分配		
			理论	实践	小计
1	第一章 数据挖掘概述	演示型	1	1	2
2	第二章 Python 数据分析与挖掘基础	演示型	1	1	2
3	第三章 认识数据	演示型	1	1	2

4	第四章 数据预处理	综合型	1	1	2
5	第五章 回归分析	综合型	1	3	4
6	第六章 分类	综合型	1	7	8
7	第七章 聚类	综合型	1	6	7
8	第八章 神经网络与深度学习	综合型	1	4	5
			8	24	32

实验类型：①演示型 ②验证型 ③设计型 ④综合型

(二) 各实验项目教学目标、内容与要求

第一章 数据挖掘概述 (2 学时)

教学目标:

1. 理解和掌握数据挖掘与机器学习的基本概念、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。
2. 了解数据挖掘与机器学习的应用和面临的问题。
3. 对数据挖掘和机器学习能够解决的问题和解决问题思路有清晰的认识。
4. 熟练应用 Jupyter notebook 的开发环境。

教学内容:

- **第一节 数据挖掘简介**
 1. 了解数据挖掘的定义和作用。
- **第二节 数据分析与数据挖掘**
 1. 熟练掌握数据分析与数据挖掘的联系。
- **第三节 数据挖掘的主要任务**
 1. 掌握数据挖掘和机器学习的主要内容。
- **第四节 数据挖掘的数据源**
 1. 了解数据挖掘用到的数据源。
- **第五节 数据挖掘使用的主要技术**
 1. 了解数据挖掘主要使用的技术，尤其是机器学习内容概述。
- **第六节 数据挖掘的主要问题**
 1. 理解数据挖掘中的主要问题。
- **第七节 数据挖掘与机器学习常用的建模工具**
 1. 了解数据挖掘与机器学习常用的商用和开源工具。
- **第八节 利用 Python 进行数据挖掘与机器学习**
 1. 掌握 Python 数据挖掘与机器学习的常用类库。
 2. 熟悉 Jupyter Notebook 的使用方法。

思想政治教育：

强调数据挖掘和机器学习技术在国家治理、社会发展中的重要性，培养学生在使用技术时的社会责任感和道德意识。

第二章 Python 数据分析与挖掘基础（2 学时）

教学目标：

1. 理解和掌握 Python 基础语法、内建的数据结构、Numpy 数值运算基础、Pandas 统计分析基础。
2. 掌握 Matplotlib 图表绘制基础等数据分析和可视化方法。

教学内容：

- **第一节 Python 程序概述**
 1. 熟练掌握 Python 编程基础。
- **第二节 Python 内建的数据结构**
 1. 熟练掌握列表、元组、字典和集合。
- **第三节 Numpy 数值计算基础**
 1. 掌握数组的变换和运算。
- **第四节 Pandas 数据分析基础**
 1. 熟练掌握 Pandas 数据分析方法。
- **第五节 Matplotlib 数据可视化基础**
 1. 熟练掌握 Matplotlib 可视化方法。
- **第六节 Scikit-learn 基础**
 1. 了解 Scikit-learn 的主要功能。
 2. 了解 Scikit-learn 的主要数据集。

思想政治教育：

强调自主创新的重要性，鼓励学生掌握开源工具，树立“科技强国”的理想和使命感。

第三章 认识数据（2 学时）

教学目标：

1. 理解和掌握数据对象和属性类型，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。
2. 了解数据可视化的方法。

教学内容：

- **第一节 数据对象与属性类型**

1. 熟练掌握数据的属性类型。
- **第二节 数据的基本统计描述**
 1. 熟练掌握基本图形显示的方法。
- **第三节 数据可视化**
 1. 了解数据可视化的方法。
- **第四节 度量数据的相似性和相异性**
 1. 熟练掌握度量数据相似性和相异性的方法。

思想政治教育：

强调数据的准确性和真实性，培养学生在数据处理中的职业道德，避免数据造假和误导性分析。

第四章 数据预处理（2 学时）

教学目标：

1. 了解数据预处理的目的是意义。
2. 掌握如何对数据进行清理。
3. 掌握如何对不同数据源的数据进行合并。
4. 掌握如何对数据进行变换，使之适合建模的需要。
5. 掌握如何对数据进行消减，使得在消减后的数据集上挖掘更有效。
6. 掌握利用 Python 进行数据预处理的方法。

教学内容：

- **第一节 数据预处理概述**
 1. 了解为什么要对数据预处理。
 2. 数据的质量要求。
- **第二节 数据清理**
 1. 理解缺失值、噪音数据的处理方法。
 2. 利用 Python 进行数据清理的方法。
- **第三节 数据集成**
 1. 理解数据集成的方法。
 2. 利用 Python 进行数据合并。
- **第四节 数据变换与数据离散化**
 1. 了解数据变换和数据离散化的方法。

思想政治教育：

强调数据隐私保护的重要性，培养学生在数据处理和分析中的伦理意识，确保个人隐私不被泄露。

第五章 回归分析 (4 学时)

教学目标:

1. 掌握回归分析原理。
2. 掌握一元线性回归分析。
3. 掌握多元线性回归分析。
4. 掌握逻辑回归。
5. 了解其他回归分析。

教学内容:

- **第一节 回归分析概述**
 1. 理解回归分析的定义与分类。
 2. 理解回归分析的过程。
- **第二节 一元线性回归分析**
 1. 掌握一元线性回归方法。
 2. 理解一元线性回归模型的参数估计。
 3. 理解一元线性回归模型的误差方差估计。
 4. 掌握利用 Python 进行一元线性回归。
- **第三节 多元线性回归**
 1. 理解多元线性回归模型。
 2. 理解多元线性模型的参数估计。
 3. 了解多元线性回归的假设检验及其评价。
 4. 掌握利用 Python 进行多元线性回归。
- **第四节 逻辑回归**
 1. 理解 Logistic 回归模型。
 2. 掌握利用 Python 进行 Logistic 回归。

思想政治教育:

强调科学求实的精神, 培养学生在分析和建模过程中的严谨态度, 避免为了追求结果而忽视数据真实性和模型合理性。

第六章 分类 (8 学时)

教学目标:

1. 了解分类及预测的基本思想、概念和意义。
2. 掌握常用的分类及预测算法 (或模型)。
3. 了解分类及预测挖掘的研究动态。
4. 掌握利用 Python 实现各种分类算法的方法。

教学内容:

- **第一节 基本概念**
 1. 了解分类和预测的定义。
- **第二节 决策树规约**
 1. 理解决策树归纳的概念。
 2. 掌握属性选择度量的方法。
 3. 理解树剪枝。
 4. 决策树的 Python 实现。
- **第三节 K 近邻算法**
 1. K 近邻算法原理。
 2. Python 算法实现。
- **第四节 支持向量机**
 1. SVM 算法原理。
 2. Python 算法实现。
- **第五节 贝叶斯分类**
 1. 理解贝叶斯定理。
 2. 掌握朴素贝叶斯分类算法。
 3. 朴素贝叶斯算法的 Python 实现。
- **第六节 模型评估与选择**
 1. 分类器性能的度量。
 2. 模型选择。
 3. 利用 Python 实现模型的评估与选择。
- **第七节 组合分类**
 1. 组合方法简介。
 2. 了解袋装、提升和 AdaBoost。
 3. 了解随机森林。

思想政治教育:

强调技术应用的社会影响，培养学生在分类和预测中考虑其对社会、经济和环境的影响，树立负责任的科学态度。

第七章 聚类 (7 学时)

教学目标:

1. 了解如何计算由各种属性和不同的类型来表示的对象之间的相异度。
2. 掌握 K-Means 聚类、层次聚类、基于密度的聚类和其他常用方法。
3. 掌握利用 sklearn 实现聚类的方法。

教学内容:

- **第一节 聚类分析的概念**
 1. 理解什么是聚类分析。
 2. 了解对聚类分析的要求。
 3. 掌握基本的聚类方法。
- **第二节 K-Means 聚类**
 1. 掌握典型的划分方法：K-Means 和 K-Medoids。
 2. 掌握利用 sklearn 实现 K-means 聚类。
- **第三节 层次方法**
 1. 理解凝聚的和分裂的层次聚类。
 2. 掌握簇间的距离度量。
 3. 分裂和凝聚的层次聚类算法。
 4. 算法实现。
- **第四节 基于密度的方法**
 1. 算法原理。
 2. 掌握 DBSCAN 算法。
 3. 掌握算法的 Python 实现。
- **第五节 其他聚类算法**
 1. 了解 STING 算法。
 2. 了解 COBWEB 算法。
 3. 了解模糊聚类算法。
- **第六节 聚类评估**
 1. 估计聚类趋势。
 2. 学习确定簇数的方法。
 3. 测定聚类质量。

思想政治教育：

强调集体协作精神，培养学生在团队合作中的沟通和协调能力，学习如何利用技术手段更好地解决实际问题。

第八章 神经网络与深度学习（5 学时）

教学目标：

1. 了解理解神经网络与深度学习的原理。
2. 掌握感知机与 BP 神经网络的原理。
3. 了解深度学习基础。

教学内容：

- **第一节 神经网络基础**
 1. 理解神经元模型。
 2. 理解感知机与多层网络。

- **第二节 BP 神经网络**
 1. 理解多层前馈神经网络。
 2. 理解后向传播算法。
 3. 了解 BP 神经网络应用。
- **第三节 深度学习**
 1. 理解深度学习基本原理。
 2. 了解常用的深度学习算法。

思想政治教育：

强调人工智能技术的社会责任，培养学生在发展和应用神经网络与深度学习技术时，考虑其对人类社会的长远影响。

(三) 教学单元对课程目标的支撑关系

课程目标 教学单元	1	2	3	4	5	6
	第一章 数据挖掘概述	√			√	
第二章 Python 数据分析与挖掘基础	√	√		√	√	√
第三章 认识数据	√	√	√	√	√	√
第四章 数据预处理	√	√	√	√	√	√
第五章 回归分析	√	√	√	√		
第六章 分类	√	√	√	√		
第七章 聚类	√	√	√	√		
第八章 神经网络与深度学习	√	√	√	√		

(三) 课程教学方法与学时分配

教学单元	教与学方式	考核方式	学时分配		
			理论	实践	小计
第一章 数据挖掘概述 (2 学时)	案例、讲授	平时表现/问题导向测试	1	1	2
第二章 Python 数据分析与挖掘基础 (2 学时)	案例、讲授、演示	平时表现/软件安装	1	1	2

第三章 认识数据 (2 学时)	案例、讲授、演示	平时表现/实际操作	1	1	2
第四章 数据预处理 (2 学时)	案例、讲授、演示	平时表现/实际操作	1	1	2
第五章 回归分析 (4 学时)	案例、讲授、实践	平时表现/实际操作	1	3	4
第六章 分类 (8 学时)	案例、讲授、实践	平时表现/实际操作	1	7	8
第七章 聚类 (7 学时)	案例、讲授、实践	平时表现/实际操作	1	6	7
第八章 神经网络 与深度学习 (5 学 时)	案例、讲授、实践	平时表现/实际操作	1	4	5
合计			8	24	32

四、课程思政教学设计

1. 在课程中讨论数据挖掘和机器学习技术对社会、经济、环境等方面的深远影响。例如，在数据挖掘概述章节，引导学生思考如何利用数据技术为国家治理、公共服务以及经济发展提供支持。
2. 强调在数据处理和分析过程中对数据隐私保护的重视，讨论机器学习模型在医疗、金融等领域的问题。
3. 鼓励学生探索前沿技术，了解国内外数据分析与机器学习的最新发展，树立自主创新意识。
4. 强调在数据分析与建模过程中的严谨性，注重科学方法的运用，反对浮躁、追求短期效应的学风。

五、课程考核

总评构成	占比	考核方式	课程目标						合计
			1	2	3	4	5	6	
1	60%	期末测试	20	20	20	20	20		100
X1	20%	平时表现	20	20	20	20	20		100
X2	10%	实验操作		25	25	25	25		100
X3	10%	实验报告		20	20	20	20	20	100

评价标准细则 (选填)

考核项	课程目	考核要求	评价标准			
			优 100-90	良 89-75	中 74-60	不及格 59-0

目	标					
1						
X1						
X2						
X3						

六、其他需要说明的问题

--