

# 《数据挖掘技术》

## 课程教学大纲

### 一、课程基本信息

课程类型	总学时为学时数	<input checked="" type="checkbox"/> 理论课（含上机、实验学时）			
	总学时为周数	<input type="checkbox"/> 实习 <input type="checkbox"/> 课程设计 <input type="checkbox"/> 毕业设计			
课程编码	7086311	总学时	48	学分	3
课程名称	数据挖掘技术				
课程英文名称	Data Mining Technology				
适用专业	统计学专业				
先修课程	(7270501) 概率论、 (7273201) 数理统计、 (7303801) Python 及其应用				
开课部门	理学院统计学系				

### 二、课程性质与目标

本课程为统计学专业选修课。本课程为学生参与学科竞赛和完成毕业论文等奠定大数据分析技术基础，目的是让学生熟悉数据的类型和预处理办法，掌握分类模型、聚类模型和关联算法，培养学生运用数据挖掘算法解决实际问题的能力。

课程目标 1：熟悉数据的类型和预处理办法，掌握决策树等分类模型、K 均值等聚类模型和 Apriori 等关联算法。

课程目标 2：运用 R 和 Python 软件编写代码，实现多种数据挖掘模型，并且利用它们解决实际问题。

课程思政目标：坚定学生理想信念、厚植爱国主义情怀、加强品德修养等，培育学生科学精神、创新精神、工匠精神等。

### 三、课程教学基本内容与要求

#### 1. 数据挖掘概述

##### 教学基本内容

- (1)数据挖掘的概念和主要任务；
- (2)本课程的主要内容和发展方向。

##### 教学基本要求

掌握数据挖掘的流程和主要任务。

## 2. 数据预处理方法

### 教学基本内容

- (1)数据的各种属性。
- (2)数据的各种预处理方法。
- (3)数据的汇总统计指标。
- (4)数据的可视化技术。
- (5)预处理或描述数据的 R 和 Python 代码。

### 教学基本要求

了解数据的属性，掌握数据的预处理方法，能够编程预处理或描述数据。

## 3. 分类模型

### 教学基本内容

- (1)分类模型概述。
- (2)决策树模型。
- (3)K-最近邻模型。
- (4)贝叶斯分类模型。
- (5)人工神经网络模型。
- (6)支持向量机模型。
- (7)集成预测模型。
- (8)各种分类模型的 R 或 Python 代码。

### 教学基本要求

掌握各种分类模型的思想，能够编程实现这些模型。

## 4. 聚类模型

### 教学基本内容

- (1)聚类分析概述。
- (2)K 均值聚类模型。
- (3)凝聚层次聚类模型。
- (4)DBSCAN 模型。
- (5)EM 算法。
- (6)各种聚类模型的 R 或 Python 代码。。

### 教学基本要求

掌握各种聚类模型的思想，能够编程实现这些模型。

## 5. 关联算法

### 教学基本内容

- (1)关联分析概述。

(2)Apriori 算法。

(3)Apriori 算法的 R 和 Python 代码。

教学基本要求

掌握 Apriori 算法，能够编程实现 Apriori 算法。

#### 6. 总复习

教学基本内容

总结本课程的主要知识点

教学基本要求

复习本课程的主要知识点

### 四、 课程学时分配

教学内容	讲授	实验	上机	课内学时小计	课外学时
1.数据挖掘概述	2			2	2
2.数据预处理方法	2			2	2
3.分类模型	20	10		30	30
4.聚类模型	4	4		8	8
5.关联算法	2	2		4	4
6.总复习	2			2	2
合 计	32	16		48	48

### 五、 实践性教学内容的安排与要求

实践性教学内容：运用 R 和 Python 软件编写代码，结合线上数据，实现决策树等分类模型，K 均值等聚类模型，以及 Apriori 等关联算法。

实践性教学场地与设备：理学院实验室（博远楼 621）或者其它小型机房，安装 R 和 Python 软件的电脑。

实践基本要求：掌握 R 和 Python 软件的常用函数命令，能够运用它们编写代码，实现决策树等分类模型，K 均值等聚类模型，以及 Apriori 等关联算法，每次课撰写一份实验报告，说明数据来源、分析过程和结果。

### 六、 教学设计与教学组织

本课程主要讲解常见的数据挖掘算法，在理论教学部分讨论模型的原理，在

实验教学部分讨论模型的代码。本课程属于专业选修课，开设在第六学期，此时，学生已经了解数据的描述方法，初步认识连续型变量的建模方法等。本课程以学生为中心，设计科学的教学方案。

1. 认真剖析教学重点和难点，综合运用多种方法，提高课堂教学质量。在讲解模型原理时，适当添加课堂练习，增加课堂互动；在讨论比较容易的知识点时，运用反转课堂转换师生角色。

2. 充分利用现代信息技术，丰富教学资源，提高学生的数据分析能力。安排学生观看网络课程，预习课程内容；运用开源软件和公开数据，开展实践教学，锻炼学生的编程能力。

3. 落实立德树人任务，添加课程思政素材，培养学生的科学精神和爱国情怀。以华为等被美国纳入实体名单为例，说明国家面临的挑战，激励学生好好学习，勇于创新，将来突破外国的科技封锁。

## 七、 教材与参考资料

### 1.教材

《数据挖掘导论》（第二版），Tan 等主编，人民邮电出版社，2020，ISBN：9787115241009。

### 2.参考资料

(1)《数据挖掘：概念与技术》（第三版），Han 等主编，机械工业出版社，2012，ISBN：9787111391401。

(2)《Python 数据分析与挖掘实战》（第二版），张良均主编，机械工业出版社，2019，ISBN：9787111640028。

(3)《R 数据分析与挖掘实战》（第一版），张良均主编，机械工业出版社，2015，ISBN：9787111516040。

## 八、 课程考核方式与成绩评定标准

本课程总评成绩由期末成绩和平时成绩构成，它们各占 50%。期末考试采用开卷、闭卷和研究报告等形式，平时成绩主要考核学生的出勤、回答问题和实验报告等情况。

## 九、 大纲制(修)订说明

无

大纲执笔人：高波

大纲审核人：徐礼文

开课系主任：刘喜波

开课学院教学副院长：李红梅

制（修）订日期：2022年2月

# 《数据挖掘技术》

## 课程实验教学大纲

### 一、 课程基本信息

课程类型	<input type="checkbox"/> 独立设置的实验课 <input checked="" type="checkbox"/> 课内实验						
课程编码	7086311	学分	3	总学时	48	实验学时	16
课程名称	数据挖掘技术						
课程英文名称	Data Mining Technology						
适用专业	统计学专业						
先修课程	(7270501) 概率论、 (7273201) 数理统计、 (7303801) Python 及其应用						
开课部门	理学院统计学系						

### 二、 课程性质与目标

见课程教学大纲

### 三、 实验的性质与任务

实验课是《数据挖掘技术》的重要组成部分，能够锻炼学生的动手能力，让他们认识数据挖掘模型的效果，促使学生运用这些模型解决他们关注的问题。

实验的任务是熟悉 R 和 Python 软件的程序结构，能够调用它们的包和函数，编写完整的代码实现数据挖掘模型；可以输入或导入数据，解决调试代码时遇到的问题；能够撰写实验报告，说明数据挖掘的过程和结果。

### 四、 实验教学内容与学时分配

序号	实验名称	学时	实验类型
1	决策树模型实验	2	综合性实验
2	贝叶斯模型实验	2	综合性实验
3	人工神经网络模型实验	2	综合性实验
4	支持向量机模型实验	2	综合性实验
5	集成分类模型实验	2	综合性实验
6	K 均值等聚类模型实验	2	综合性实验

7	期望最大模型实验	2	综合性实验
8	Apriori 算法实验	2	综合性实验

## 五、 实验安排与要求

实验课安排在模型的理论课后，理学院实验室或者其它小型机房。学生可以使用实验室的电脑，也可以携带自己的电脑。

实验要求如下：

- (1) 爱护公共财产，遵守安全制度，树立优良学风；
- (2) 下载、安装 R 或 Rstudio 等编写 R 代码的软件，Python、Anaconda、Pytorch 等编写 Python 代码的软件；
- (3) 运用两种软件编写代码，完成每次课的数据挖掘任务；
- (4) 撰写实验报告，说明研究问题、数据来源、数据挖掘过程和结果等。

## 六、 实验教学与其它相关课程的联系与分工

实验教学是《数据挖掘技术》的重要组成部分。《R 程序语言》和《Python 及其应用》分别讲述 R 程序和 Python 程序进行统计分析的基本方法，为本课程的实验教学提供必需的知识基础。

## 七、 实验教学设计与教学组织

本课程以学生为中心，设计科学的教学方案。

1. 要求学生通过网络课程等预习课程内容。
2. 每节课都重点讲解若干包和函数；巡查学生的代码调试结果，共同解决他们遇到的问题。
3. 鼓励学生运用所学知识研究热点问题，培养他们的科学精神和爱国情怀。

## 八、 实验教材、实验指导书及教学参考资料

### 1. 实验教材

《Python 数据分析与挖掘实战》（第二版），张良均主编，机械工业出版社，2019，ISBN：9787111640028。

### 2. 实验指导书

包括实验任务、实验步骤、实验报告等。

### 3. 参考资料

《R 数据分析与挖掘实战》（第一版），张良均主编，机械工业出版社，，2015，ISBN：9787111516040。

## 九、 实验考核方法及成绩评定标准

实验考核是本课程平时成绩的重要组成部分，取决于学生在每次课的出勤、实验操作和实验报告等情况。

## 十、 大纲制(修)订说明

无

大纲执笔人：高波

大纲审核人：徐礼文

开课系主任：刘喜波

开课学院教学副院长：李红梅

制（修）订日期：2022年2月