课程详述
COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	数据挖掘及商务应用 Data Mining and Business Applications
2.	授课院系 Originating Department	信息系统与管理工程系 Division of Information Systems & Management Engineering
3.	课程编号 Course Code	MIS306
4.	课程学分 Credit Value	3
5.	课程类别 Course Type	专业核心课 Major Core Courses
6.	授课学期 Semester	秋季 Fall
7.	授课语言 Teaching Language	英语 English
8.	授课教师、所属学系、联系方式（如属团队授课，请列明其他授课教师） Instructor(s), Affiliation& Contact （For team teaching, please list all instructors）	王松昊，信息系统与管理工程系， wangsh2021@sustech.edu.cn Wang Songhao, Division of Information Systems & Management Engineering, wangsh2021@sustech.edu.cn
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	待公布 To be announced
10.	选课人数限额(可不填) Maximum Enrolment（Optional）	40

11.	授课方式 Delivery Method	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other（Please specify）	总学时 Total
	学时数 Credit Hours	32		32		64

12.	先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	MIS205 数据管理与数据库或 EBA203 管理信息系统
13.	后续课程、其它学习规划 Courses for which this course is a pre-requisite	MIS402 大数据实证研究
14.	其它要求修读本课程的学系 Cross-listing Dept.	无 None

教学大纲及教学日历 SYLLABUS

15.   教学目标 Course Objectives

使学生了解数据挖掘技术，并利用这些技术进行适当的数据分析，从而获得满足业务目标/策略的最佳知识库

为学生提供扎实的统计基础，以帮助他们解决问题

准备让学生理解数据分析的结果，并从业务角度应用这些结果

To provide students with an understanding of data mining techniques and use of these techniques to perform appropriate data analysis in deriving an optimal knowledge base in meeting business goals/strategies

To equip the students with a strong statistical foundation in which to employ in solving problems

Prepare the students to understand the results from the data analysis and apply those results from a business perspective

16.   预达学习成果 Learning Outcomes

学生将能够证明有关数据挖掘概念，过程的知识，并使用数据挖掘和统计技术进行数据分析

学生将能够解释数据分析的结果以及如何将这些结果应用于业务问题，目标和组织策略

Students will be able to demonstrate knowledge of data mining concepts, processes, and perform data analysis using data mining and statistical techniques

Students will be able to interpret results from data analysis and how to apply those results to business problems, goals, and organizational strategies

17.   课程内容及教学日历 （如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）
Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

理论（32 学时）

第 1 周 数据挖掘介绍（2 学时）

本次课程介绍什么是数据挖掘，包含数据挖掘的主要任务、被挖掘的有用信息、常用的技术手段及应用等。

第 2 周 数据预处理（2 学时）

本次课程介绍与数据相关的一些问题，包含数据类型、数据质量及数据预处理的常用技术等。

第 3-5 周 分类：概念与常用模型 （6 学时）

本次课程主要介绍分类的基本概念和常用技术，包含逻辑回归、决策树、临近分类、贝叶斯分类、支持向量机、集成学习等模型和方法。

理论课 6-7（4 学时）：回归：概念、常用线性及非线性回归模型

本次课程主要介绍回归的基本概念和常用技术，包含线性回归、多项式回归、岭回归、神经网络、RBF 模型、高斯过程回归模型等。

理论课 8-9（4 学时）：关联分析：概念及常用技术

本次课程主要介绍关联分析的概念及常用技术，包含频繁项集产生、规则产生、关联模式评估等。

理论课 10-11（4 学时）：聚类分析：概念及常用算法

本次课程介绍聚类的概念和常用算法，包含 k-均值、层次聚类、基于密度的聚类、簇评估等方法。

理论课 12（2 学时）：异常检测

本次课程介绍异常检测的概念和方法，包含基于统计、临近度、密度、聚类等的方法。

理论课 13（2 学时）：文本挖掘简介

本次课程简要介绍文本挖掘的一些概念，包含词频、词袋模型、词向量模型、词频-逆文档频率、N-Gram 模型等。

理论课 14（2 学时）：模型评估

本次课程结合之前课程的相关内容，介绍多种数据挖掘模型的评估方法。

理论课 15（2 学时）：数据挖掘的商务应用

本次课程通过若干实例介绍数据挖掘技术在商务领域的应用。

理论课 16（2 学时）：课程总结及数据挖掘发展方向

本次课程总结本门课所授内容并简要介绍数据挖掘的发展方向。

实验（32 学时）

第一周 Python 简介（2 学时）

本次课程介绍 python 的基础以及基于 python 的数据挖掘平台。

第二周 数据预处理上机实验（2 学时）

本次课程介绍数据描述、分析、预处理等实现方法。

第 3-5 周 数据分类上机实验（6 学时）

本次课程介绍包含逻辑回归、决策树等常用数据分类的实现方法。

第 6-7 周 数据回归上机实验（4 学时）

本次课程介绍常用的线性及非线性数据回归的实现方法。

第 8-9 周 关联分析上机实验（4 学时）

本次课程介绍关联分析的实现方法及应用。

第 10-11 周 聚类分析上机实验（4 学时）

本次课程介绍包含 k-means、层次聚类等常用聚类分析的实现方法。

第 12 周 异常检测上机实验（2 学时）

本次课程介绍常用异常检测算法的实现方法。

第 13 周 文本挖掘上机实验（2 学时）

本次课程简要介绍常用文本挖掘算法的实现方法和应用。

第 14 周 课程项目上机实验（1）（2 学时）

本次课程指导学生小组课程项目。

第 15 周 课程项目上机实验（2）（2 学时）

本次课程指导学生小组课程项目。

第 16 周 课程项目汇报（2 学时）

本次课程学生完成小组课程项目汇报。

Lecture (32 credit hours)

Week 1 Introduction to data mining (2 credit hours)

This lecture gives a general introduction of data mining, including its main tasks, the knowledge to be mined, the techniques adopted and the target applications.

Week 2 Data preprocessing (2 credit hours)

This lecture covers several topics related to data, including data type, data quality and techniques for data preprocessing, etc.

Week 3-5 Classification: concepts and techniques for classification (6 credit hours)

These lectures discuss the fundamental concepts and techniques for classification, including logistic regression, decision tree, k-nearest-neighbor classifiers, Bayesian classifiers, support vector machine, ensemble methods, etc.

Week 6-7 Regression: concepts, linear & non-linear regression models (4 credit hours)

These lectures discuss the concepts and models for regression, including linear regression, polynomial regression,

ridge regression, neural networks, radial basis function (RBF) models, Gaussian process regression, etc.

Week 8-9 Association analysis: concepts and techniques (4 credit hours)

These lectures introduce the concepts and techniques for association analysis, including frequent itemset generation, rule generation, pattern evaluation methods, etc.

Week 10-11 Cluster analysis: concepts and algorithms (4 credit hours)

These lectures introduce the concepts and algorithms for cluster analysis, including K-means, hierarchical clustering, density-based clustering, cluster evaluation, etc.

Week 12 Outliers detection (2 credit hours)

This lecture introduces concepts and techniques for outlier detection, including statistical approaches, proximity-based approaches, density-based approaches, clustering-based approaches, etc.

Week 13 Introduction to text mining (2 credit hours)

This lecture briefly introduces the concepts of text mining, including term frequency, bag of words, word embedding, TFIDF, N-Gram model, etc.

Week 14 Model evaluation (2 credit hours)

This lecture introduces several methods to evaluate different data mining models.

Week 15 Data mining in business (2 credit hours)

This lecture discusses several examples and case studies of data mining applications in business.

Week 16 Review and data mining trends (2 credit hours)

This lecture reviews the topics in this module and presents some trends of data mining.

Lab (32 credit hours)

Week 1. Introduction to Python (2 credit hours)

This lecture introduces the basic knowledge of Python and the data mining platform based on python.

Week 2. Lab tutorial for data preprocessing (2 credit hours)

This lecture introduces the techniques for data description, analysis and preprocessing.

Week 3-5. Lab tutorial for classification techniques (6 credit hours)

These lectures introduce the implementations of common classification algorithms including logistic regression, decision tree, etc.

Week 6-7. Lab tutorial for regression models (4 credit hours)

These lectures introduce the implementations of common linear & non-linear regression algorithms.

Week 8-9. Lab tutorial for association analysis (4 credit hours)

These lecture introduces the implementations and applications of association analysis techniques.

Week 10-11. Lab tutorial for cluster analysis (4 credit hours)

These lectures introduce the implementations of common cluster analysis algorithms including k-means, hierarchical clustering, etc.

Week 12. Lab tutorial for outliers detection (2 credit hours)

This lecture introduces the implementations of common algorithms for outliers detection.

Week 13. Lab tutorial for text mining (2 credit hours)

This lecture briefly introduces the implementations of texting mining algorithms as well as their applications.

Week 14. Lab tutorial for group project I (2 credit hours)

This tutorial guides the students for the group projects.

Week 15. Lab tutorial for group project II (2 credit hours)

This tutorial guides the students for the group projects.

Week 16. Project presentation (2 credit hours)

This tutorial is for the group project presentation.

18. 教材及其它参考资料 Textbook and Supplementary Readings

Introduction to Data Mining (Second Edition), by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar

Data Mining: Concepts and Techniques (Third Edition), by Jiawei Han, Micheline Kamber, and Jian Pei

Data Science for Business, by Foster Provost, and Tom Fawcett

机器学习，周志华著

课程评估 ASSESSMENT

19.

评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance		5		
课堂表现 Class Performance				
小测验 Quiz				
课程项目 Projects		15		上机实验
平时作业 Assignments		15		
期中考试 Mid-Term Test				
期末考试 Final Exam		40		
期末报告 Final Presentation		25		分组项目汇报
其它（可根据需要改写以上评估方式） Others (The above may be modified as necessary)				

20. 记分方式 GRADING SYSTEM

☑ A. 十三级等级制 Letter Grading
☐ B. 二级记分制（通过/不通过） Pass/Fail Grading

课程审批 REVIEW AND APPROVAL

21. 本课程设置已经过以下责任人/委员会审议通过
This Course has been approved by the following person or committee of authority