

MIS303 课程大纲

- 1、2022 秋季学期 (2-7页码)
- 2、2023 秋季学期起 (8-15页码)

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	大数据治理与商业模式 Big Data Governance and Business Model
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	MIS303
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)	黄伟, 信息系统与管理工程系, huangw7@sustech.edu.cn Wei Huang, Department of Information Systems & Management Engineering
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)	

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	管理信息系统、数据管理与数据库 Management Information Systems, Data Management and Database				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	无 None				
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

课程旨在培养学生将数据治理融合到机构的商业模式中。学生将学会如何运用信息技术，根据法规和机构运营制定并执行规章制度以保护组织的信息资产。通过介绍制定有效数据治理的流程，课程将帮助学生掌握如何将来自企业各个部门的不同期望和专业领域汇集在一起，从而做出有数据支持的决策。同时，课程将介绍有效数据治理的完整生命周期，即从元数据管理到隐私和合规。

This course intends to provide students with knowledge of how data governance fits into an organization's business model. Students will learn how to work with information technology, regulations and operations to set policies and initiate good practices that can safeguard an organization's informational assets. By introducing the processes for developing an effective data governance, this course intends to enhance students' understanding of how an organization can bring together diverse expectations and expertise across the enterprise to enable data-informed decision making. In addition, this course will introduce students to the complete life cycle of effective data governance from metadata management to privacy and compliance.

16. 预达学习成果 Learning Outcomes

At the end of this course, students will be able to

- understand what data are and how they work 了解什么是数据以及数据的运作
- assess the business issues that data management can resolve 评估数据管理能够解决的业务问题
- evaluate and explain the challenges inherent in data management and governance 评估数据管理和治理遇到的挑战
- explain data governance maturity models 解释完整的数据治理模型
- understand how to mitigate regulatory and operational risk through data governance 掌握如何通过数据治理减轻监管和业务风险

- articulate a data quality strategy 阐明数据质量策略
- document the business needs for data governance 记录数据治理的商务业务需求
- create a data quality improvement plan 创建数据质量改进计划.

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.)

This course will provide an overview of data governance and business model. It will cover building a governance infrastructure based on business model, agents' roles and responsibilities, stewardship, governance communications, regulatory compliance, privacy concerns, data ethics, and risk management. This course will address data quality as a continuous issue in data management. It will emphasize the challenges of data quality in the context of big data as volumes of data increase and the uses for data expand. This course will utilize case studies, trends, techniques, and best practices as it examines the topics of data governance and business model.

Topics include the following.

Chapter 1 Data Concepts and Definitions (4 hours)

Chapter 2 Data Mining and Data Analysis Techniques (4 hours)

Chapter 3 Data Governance Principles and Risk (2 hours)

Chapter 4 Metadata Management and Data Governance (2 hours)

Chapter 5 Building a Business Case for Data Governance (4 hours)

Chapter 6 Data Quality Through Governance (4 hours)

Chapter 7 Data Quality Assessment and Measurement (2 hours)

Chapter 8 Data Assessment Scenarios (2 hours)

Chapter 9 A Strategic Approach to Data Quality (4 hours)

Chapter 10 The Future of Data Governance (4 hours)

The lab include:

Chapter 1 Description Analysis (2 hours)

1.1 Data analysis (1 hours)

This section mainly shows how to use the "data analysis" function in Excel to analyze data.

1.2 Pivot Table (1 hours)

This section mainly shows how to use the "Pivot Table" to visually analyze the data.

Chapter 2 Python & R (2 hours)

This chapter focuses on Python & R installation and the basic syntax of Python & R.

2.1 Python & R installation

2.2 Basic syntax of Python & R

Chapter 3 Web Crawler (6 hours)

3.1 Requests library (2 hours)

This section explains the HTML tags and the use of Requests library in Python.

3.2 BeautifulSoup Library - Picture Grab (2 hours)

This section focuses on the BeautifulSoup library in Python and uses of it to grab images from internet.

3.3 BeautifulSoup Library - Text Grab (2 hours)

This section is mainly to tutors students program web crawler to grab text information from the internet.

Chapter 4 Text Mining (4 hours)

4.1 jieba library (2 hours)

This section mainly explains the processing of jieba library.

4.2 TF-IDF algorithm (2 hours)

This section mainly explains how to use the TF-IDF algorithm to extract keywords and seek text similarity.

Chapter 5 Network Data Processing (4 hours)

2.1 SQL statement (2 class hours)

This part mainly explains the basic SQL statements: select, update, insert, delete

2.2 Generation of Network Data (2 class hours)

This part mainly explains how to use SQL statements to clean data and generate social network data that can be used for analysis.

Chapter 6 Network Visualization (4 hours)

3.1 NetDraw (2 hours)

This part mainly provides an overview of Netdraw software and explains the structure of VNA data.

3.2 Data Visualization (2 hours)

This section explains how to use Netdraw software for social network data visualization analysis.

Chapter 7 Data Management Cases (4 hours)

Case Study on Data quality

Chapter 8 Course Project (6 hours)

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

Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program (The Morgan Kaufmann Series on Business Intelligence) 1st Edition

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	课上			
课堂表现 Class Performance	课上	10		
小测验 Quiz	课上			
课程项目 Projects	学期中	20		
平时作业 Assignments				
期中考试 Mid-Term Test				
期末考试 Final Exam	期末	30		
期末报告	期末	40		

Final Presentation

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

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	大数据治理与商业模式 Big Data Governance and Business Model
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	MIS303
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	中英文 Chinese & 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)	胡大宁, 信息系统与管理工程系, hudn@sustech.edu.cn HuDaning, Department of Information Systems & Management Engineering
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)	

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	管理信息系统、数据管理与数据库 Management Information Systems, Data Management and Database				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	无 None				
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

课程旨在培养学生将数据治理融合到机构的商业模式中。学生将学会如何运用信息技术，根据法规和机构运营制定并执行规章制度以保护组织的信息资产，并学习使用因果推断的框架来理解和分析企业所积累的大量二手数据。通过介绍制定有效数据治理的流程，课程将帮助学生掌握如何将来自企业各个部门的不同期望和专业知识汇集在一起，从而做出有数据支持的决策。同时，课程将介绍有效数据治理的完整生命周期，即从元数据管理到隐私和合规。

This course intends to provide students with knowledge of how data governance fits into an organization's business model. Students will learn how to work with information technology, regulations and operations to set policies and initiate good practices that can safeguard an organization's informational assets, as well as how to use causal inference methods to understand and analyze large scale secondary data accumulated by firms' daily operations. By introducing the processes for developing an effective data governance, this course intends to enhance students' understanding of how an organization can bring together diverse expectations and expertise across the enterprise to enable data-informed decision making. In addition, this course will introduce students to the complete life cycle of effective data governance from metadata management to privacy and compliance.

16. 预达学习成果 Learning Outcomes

At the end of this course, students will be able to

- understand what data are and how they work 了解什么是数据以及数据的运作
- assess the business issues that data management can resolve 评估数据管理能够解决的业务问题
- evaluate and explain the challenges inherent in data management and governance 评估数据管理和治理遇到的挑战
- learn causal inference data analysis methods from empirical research 学习因果推断框架来理解和分析企业大数据
- explain data governance maturity models 解释完整的数据治理模型
- understand how to mitigate regulatory and operational risk through data governance 掌握如何通过数据治理减轻监管和业务风险



- articulate a data quality strategy 阐明数据质量策略
- document the business needs for data governance 记录数据治理的商务业务需求
- create a data quality improvement plan 创建数据质量改进计划.

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.)



这门课程将概述数据治理和商业模型。它将涵盖基于商业模型构建治理基础设施、代理人的角色和责任、管家职责、治理沟通、法规遵守、隐私问题、数据伦理和风险管理。本课程将把数据质量作为数据管理中的一个持续问题进行讲解。它将强调在大数据背景下数据质量的挑战，因为数据量的增加和数据用途的扩展。本课程还将教授因果推断方法，以分析企业中的大规模二手数据。本课程将利用案例研究、趋势、技术和最佳实践，深入研究数据治理和商业模型主题。

主题包括以下几个方面。

第1章 数据概念与定义（4小时）

第2章 数据挖掘与数据分析技术（4小时）

第3章 数据治理原则与风险（2小时）

第4章 元数据管理与数据治理（2小时）

第5章 为数据治理构建商业案例（4小时）

第6章 通过治理实现数据质量（4小时）

第7章 数据质量评估与测量（2小时）

第8章 数据评估场景（2小时）

第9章 数据质量的战略方法（4小时）

第10章 数据治理的未来（2小时）

第11章 使用因果推断理解和分析企业中的大规模二手数据（4小时）

实验室包括：

第一章 描述分析（2小时）

1.1 数据分析（1小时）

本节主要介绍如何在Excel中使用“数据分析”功能来分析数据。1.2 数据透视表（1小时）

本节主要介绍如何使用“数据透视表”进行数据的可视化分析。

第二章 Python和R语言（2小时）

本章重点介绍Python和R语言的安装以及Python和R语言的基本语法。

2.1 Python和R语言的安装

2.2 Python和R语言的基本语法

第三章 网络爬虫（6小时）

3.1 Requests库（2小时）

本节解释HTML标签以及Python中Requests库的使用。

3.2 BeautifulSoup库 - 图片抓取 (2小时)

本节关注Python中的BeautifulSoup库以及如何使用它从互联网抓取图片。3.3 BeautifulSoup库 - 文本抓取 (2小时)

本节主要教导学生如何编程使用网络爬虫从互联网抓取文本信息。

第四章 文本挖掘 (4小时)

4.1 jieba库 (2小时)

本节主要讲解jieba库的处理方法。

4.2 TF-IDF算法 (2小时)

本节主要解释如何使用TF-IDF算法提取关键词和寻求文本相似性。

第五章 网络数据处理 (4小时)

2.1 SQL语句 (2课时)

本部分主要讲解基本的SQL语句: select, update, insert, delete

2.2 网络数据生成 (2课时)

本部分主要讲解如何使用SQL语句清洗数据并生成可用于分析的社交网络数据。

第六章 网络可视化 (4小时)

3.1 NetDraw (2小时)

本部分主要提供Netdraw软件的概述, 并解释VNA数据的结构。3.2 数据可视化 (2小时)

本节解释如何使用Netdraw软件进行社交网络数据可视化分析。

第七章 数据管理案例 (4小时)

数据质量案例研究

第八章 课程项目 (6小时)

This course will provide an overview of data governance and business model. It will cover building a governance infrastructure based on business model, agents' roles and responsibilities, stewardship, governance communications, regulatory compliance, privacy concerns, data ethics, and risk management. This course will address data quality as a continuous issue in data management. It will emphasize the challenges of data quality in the context of big data as volumes of data increase and the uses for data expand. This course will also teach causal inference methods to analyze large scale secondary data in firms. This course will utilize case studies, trends, techniques, and best practices as it examines the topics of data governance and business model.

Topics include the following.

Chapter 1 Data Concepts and Definitions (4 hours)

Chapter 2 Data Mining and Data Analysis Techniques (4 hours)

Chapter 3 Data Governance Principles and Risk (2 hours)

Chapter 4 Metadata Management and Data Governance (2 hours)

Chapter 5 Building a Business Case for Data Governance (4 hours)

Chapter 6 Data Quality Through Governance (4 hours)

Chapter 7 Data Quality Assessment and Measurement (2 hours)

Chapter 8 Data Assessment Scenarios (2 hours)

Chapter 9 A Strategic Approach to Data Quality (4 hours)

Chapter 10 The Future of Data Governance (2 hours)

Chapter 11 Using Causal Inference to understand and analyze large scale secondary data in firms (4 hours)

The labs include:

Chapter 1 Description Analysis (2 hours)

1.1 Data analysis (1 hours)

This section mainly shows how to use the "data analysis" function in Excel to analyze data.

1.2 Pivot Table (1 hours)

This section mainly shows how to use the "Pivot Table" to visually analyze the data.

Chapter 2 Python & R (2 hours)

This chapter focuses on Python & R installation and the basic syntax of Python & R.

2.1 Python & R installation

2.2 Basic syntax of Python & R

Chapter 3 Web Crawler (6 hours)

3.1 Requests library (2 hours)

This section explains the HTML tags and the use of Requests library in Python.

3.2 BeautifulSoup Library - Picture Grab (2 hours)

This section focuses on the BeautifulSoup library in Python and uses of it to grab images from internet.

3.3 BeautifulSoup Library - Text Grab (2 hours)

This section is mainly to tutors students program web crawler to grab text information from the internet.

Chapter 4 Text Mining (4 hours)

4.1 jieba library (2 hours)

This section mainly explains the processing of jieba library.

4.2 TF-IDF algorithm (2 hours)

This section mainly explains how to use the TF-IDF algorithm to extract keywords and seek text similarity.

Chapter 5 Network Data Processing (4 hours)

2.1 SQL statement (2 class hours)

This part mainly explains the basic SQL statements: select, update, insert, delete

2.2 Generation of Network Data (2 class hours)

This part mainly explains how to use SQL statements to clean data and generate social network data that can be used for analysis.

Chapter 6 Network Visualization (4 hours)

3.1 NetDraw (2 hours)

This part mainly provides an overview of Netdraw software and explains the structure of VNA data.

3.2 Data Visualization (2 hours)

This section explains how to use Netdraw software for social network data visualization analysis.

Chapter 7 Data Management Cases (4 hours)

Case Study on Data quality

Chapter 8 Course Project (6 hours)

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

Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program (The Morgan Kaufmann Series on Business Intelligence) 1st Edition

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	课上			
课堂表现 Class Performance	课上	10		
小测验 Quiz	课上			
课程项目 Projects	学期中	20		
平时作业 Assignments				
期中考试 Mid-Term Test				
期末考试 Final Exam	期末	30		
期末报告	期末	40		



Final Presentation				
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

胡大宇 陆晔