



概 述 (课程简介)

徐华

xuhua@tsinghua.edu.cn

1

基本信息



◎ 教师与助教

- ◆ 教师：徐华，xuhua@tsinghua.edu.cn，13810102923 (M)
- ◆ 助教：吴至婧，wuzj18@mails.tsinghua.edu.cn，15528148260(M)
- ◆ Office Hour：每周一中午1:30-2:30（需要提前预约，FIT 4-506房间）

◎ 上课时间

- ◆ 周一第六大节(19:20 开始)
- ◆ 除法定节假日外，全周

◎ 上课地点 新水325

◎ 先修课程要求及预备知识

- ◆ 英语基础
- ◆ 基本的程序设计语言(Python, R, Java, C#, C++, C 等不限)
- ◆ 一定的高等数学背景知识

◎ 交流及讨论：清华网络学堂、课程微信群

2



课程定位



课程对象

- ◆ 全校 **非信息类 理工科专业** 的本科生

课程目的

- ◆ 具有数据挖掘与分析的**基本思想方法**。
- ◆ 形成**大数据思维方法**。

课程定位

- ◆ 掌握思路，了解技术，解决问题，跟踪应用，挑战提高。
- ◆ 针对不同的工业技术领域，可独立开展相应的课程实践。

课程学时：48学时 / 3 学分

- ◆ 课堂讲授（课堂讲授 + 高级专题(专题讨论) + 应用案例讲解）
- ◆ 课程实践（2个课程作业(数据获取作业 + 算法实验作业) + 1个具有展示度的软件/程序（共享）或者 **参加挑战性数据挖掘任务（难度较大）**）
- ◆ 课程考试（有限开卷，第14周，**2021年5月24日**）

3

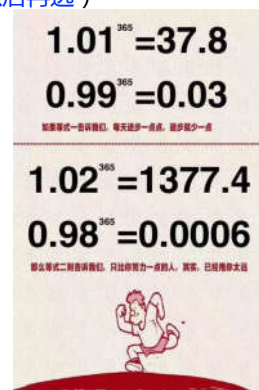


课程定位



几点要求

- ◆ 对数据挖掘与分析工作感兴趣
- ◆ 未来希望从事领域或者行业数据挖掘与分析工作
- ◆ 本学期能够投入一定的学习时间（**学习时间无法保障的同学建议以后再选**）



4



教学内容及考核(1)



● 教学目标

- ◆ 建立对数据挖掘的总体认识，学会从数据挖掘的角度来思考、分析和解决问题
- ◆ 掌握数据挖掘的基本概念与方法，懂得如何系统化地解决实际问题

● 教学形式：双语教学（英文课件，中文讲解，推荐双语参考书）

● 教学计划：15次课（按照校历，第11周停课1次）

- ◆ 基本方法与原理讲解：10 次课
- ◆ 应用与专题讲解：2次课
- ◆ 中后期课程考试：1 次课（第14周 周一 5月23日）
- ◆ 课程实践讨论课：2 次课
- ◆ 具体教学计划安排如下表

5



教学内容及考核(2)



Topic	#wk
Introduction	1
Data Pre-processing	2
Data Warehouse	1
Association and Correlation Analysis	2
Classification and Prediction	2
Cluster Analysis	2
One Application Example Introduction	1
Advanced Topic	1
Experiment Presentation and Discussion	1+1
Course Test	1
	15

6



教学内容及考核(3)



- ◎ **教学考核**：最终成绩 = 课程作业(自选)*60% + 课程考试(独立有限开卷)*40%
- ◎ **作业要求**：报告/共享
 - ◆ 课程3次小作业（参加**挑战任务**的同学不参加）
 - 作业1：特定领域的获取数据，获取结果上传课程共享GitHub上实时更新（同组可上传多次）
 - 作业2：在已经获取的数据集上，开展一个数据挖掘分析与实验工作
 - 作业3：在已经获取的数据集上进行一定的数据可视化呈现
 - ◆ 课程大作业建议（**二选一**）：
 - **挑战任务**：挑战性数据挖掘任务（由企业导师指导,课程教学团队提供必要的支持）
 - 自选任务：在自有数据集或者教师分享的数据集上完成挖掘任务与应用汇报（个人为单位）
- ◎ **课程实验**：选题/讨论/报告/展示
 - ◆ 第6-13周期间，小作业的抽查与汇报，或者相关研究工作分享
 - ◆ 第15周之前，大作业软件/程序在网络学堂平台上面向全体同学开放共享
 - ◆ 第15-16周，大作业课堂展示与抽查汇报

7



教学内容及考核(3)



- ◎ **课程考试**：14周（**5月23日，周一**），独立有限开卷测验
- ◎ **关于课程成绩**：
 - ◆ 课程成绩：等级成绩（以往不通过的情况：未完成作业 或 未参加考试者）。
 - ◆ 课程特区：数据挖掘挑战任务

8





概述结束!

