

## 课程教学进度计划表

## 一、基本信息

课程名称	数据分析与挖掘				
课程代码	2050544	课程序号	4729	课程学分/学时	48
授课教师	董辛酉	教师工号	20048	专/兼职	专
上课班级	B21-67	班级人数	25	上课教室	信息 412、三教 108
答疑安排	周二 5-8 节, 面授和线上				
课程号/课程网站					
选用教材	【Python 网络爬虫技术与实践, 吕云翔著, 机械工业出版社, 2023 年 6 月】				
参考教材与资料	【Python 程序设计 (第 2 版), 董付国, 清华大学出版社, 2016. 6】				

## 二、课程教学进度安排

课次	课时	教学内容	教学方式	作业
1	2	Python 环境配置; 掌握基础语法学习;	线下教学	
2	2	网络爬虫原理和技术讲解;	边讲边练	1
3	2	掌握数据采集方法, 如开放数据集、API (应用程序接口)、爬虫技术和传感器数据;	边讲边练	1
4	2	设计并实现可存储金融数据的数据库;	边讲边练	1
5	2	掌握数据预处理方法, 如缺失值处理、异常值处理、数据转换、数据合并与重塑等;	边讲边练	1
6	2	掌握数据预处理方法, 特征工程、数据分割、时间序列处理等;	边讲边练	1
7	2	网络爬虫技术实战;	边讲边练	1
8	2	掌握常用的静态网页采集技术: 1) HTTP 请求; 2) HTML 解析;	线下教学	

9	2	了解基于 Tornado 的分布式 Web 框架；	线下教学	
10	2	掌握大数据存储方式；	边讲边练	1
11	2	介绍 Cassandra、MongoDB 等非关系型数据库；	边讲边练	1
12	2	了解 Kafka 分布式流处理平台；	边讲边练	1
13	2	在分布式系统中，添加日志功能；	线下教学	1
14	2	掌握网页动态内容的获取方式； 掌握 DOM（文档对象模型）；	线下教学	1
15	2	掌握爬虫模拟用户行为的方式； 掌握金融类数据的爬取方式；	边讲边练	1
16	2	掌握模拟登陆和验证码处理的原理和技术；	边讲边练	1
17	2	实现金融类数据的爬取流程；	边讲边练	1
18	2	实现数据的预处理和清洗；	边讲边练	1
19	2	掌握数据分析和可视化相关技术；	边讲边练	1
20	2	掌握 TA-Lib 库；	边讲边练	1
21	2	实现 TA-Lib 的相关技术指标和形态的计算；	线下教学	
22	2	完善并总结待评审系统；	线下教学	
23	2	课堂答疑、大作业评审	线下教学	
24	2	大作业评审	线下教学	

### 三、考核方式

总评构成	占比	考核方式
X1	大作业	40%
X2	实验报告	30%
X3	课堂表现	30%

任课教师：董辛酉 系主任审核：戴志明

日期：20240830