《Python与城市大数据分析》教学大纲

课程编号：

课程类型：□通识教育必修课 □通识教育选修课

□学科基础课 □√专业选修课

总 学 时：32 讲课学时：16 实验（上机）学时：16

学　　分：2

考试类型：□考试 □√考查

适用对象：城市管理专业

□√是 □否 适合作为其他专业学生的个性化选修课

先修课程：计算机应用

一、教学目标

大数据已经成为城市活动的重要基础和各学科关注的重点，但传统的数据处理技术和工具不太适用于大数据的情形，新兴的处理技术和工具迫切需要引入学生课堂。Python语言作为处理大数据的成熟工具，已被广泛应用于学术界和实业当中。本课程的目的是讲述Python语言的基本语法及其在城市大数据分析中的基本应用，侧重于文本分析、数据处理、网络爬虫和数据可视化等代表性的大数据处理技术。通过本课程的学习，使学生了解大数据的采集、传输、处理、分析和应用的相关技术，掌握Python的基础语言知识、大数据分析和应用技术，能够基于Python进行数据采集、处理、分析和应用开发工作，为未来进一步深入学习与实践奠定基础。

课程思政教学目标：通过城市大数据分析案例，开展行业规范教育、加强学生学术道德、职业操守与法律意识。

二、教学内容及其与毕业要求的对应关系

（一）教学内容

本课程主要教学内容是在学生掌握计算机应用与大数据分析基本原理后，学习利用Python语言编程，掌握利用计算机解决问题的基本方法和过程。具体包括： Python基本语法函数、数据读取与存储、文本分析、网络爬虫、数据处理与可视化、案例分析等。主要介绍或项目使用到的Python第三方库有：jieba分词，Requests，BeautifulSoup，Re，Scrapy, Pandas，Pyecharts, Matplotlib等。

（二）教学方法和手段

根据教学目标，拟采用的教学方法有：课堂讲解基本概念和核心知识，讲授和讨论相结合领会知识要点，案例教学训练解决问题的能力，借助Python编程软件让学生进行上机操作和具体实践。

（三）实践教学环节要求

根据教学进度和要求布置课堂上机操作和课后实践小作业。由于课程实操较强，建议学生安装和配置软件进行练习，完成案例的内容。通过上机学习本课程的理论，掌握Python的基础语言知识，能够基于Python进行数据采集、处理、分析和应用开发工作。

（四）学习要求

Python作为一门高级编程语言适合零基础人员学习，为有效学习本课程，要求学生首先熟悉程序设计基础与应用、计算机网络技术与应用、数据库等基本原理。通过本课程的学习掌握Python语言编程的基本知识和技能，具体包括：Python的数字类型、字符串类型、分支、循环、函数、列表类型、字典类型、文件和第三方库使用等概念和编程方法。掌握文本分析、定向网络数据爬取和网页解析、数据处理与可视化的基本能力。通过课上上机练习和课后思考进行多角度和多层次的反复学习。

（五）与毕业要求的关系

随着大数据时代的到来，数字化、智慧化城市管理是未来发展的必然趋势。因此，学习和掌握城市大数据的分析方法、技术及相关应用，是当前城市智慧化管理的需要，也是高校城市管理专业培养的基本要求。课程内容是进入大数据处理、数据挖掘、以数据为中心人工智能领域的必备实践基础。

（六）教学中应注意的问题

由于程序设计基础与应用、计算机网络技术与应用等是该课程的基础，如果缺少上述基础，学习Python语言的编程及数据分析工具，容易出现不理解编程数据分析结果的问题。因此，教学中需要根据学生掌握先修课程基础情况，由易到难循序渐进学习编程及数据分析相关工具和技术；并且指导学生在上课过程中根据课程进度补充相关知识并多进行编程练习。

三、各教学环节学时分配

各章节的学时分配如下：

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 章节内容 | 讲课 | 实验 | 其他 | 合计 |
| 1 | 大数据概念与Python基本语法 | **2** | **2** |  | **4** |
| 2 | 条件语句、循环与函数 | **2** | **2** |  | **4** |
| 3 | 数据读取与存储 | **2** | **2** |  | **4** |
| 4 | 文本分析初步 | **2** | **2** |  | **4** |
| 5 | 网络爬虫入门 | **2** | **4** |  | **6** |
| 6 | 数据处理 | **2** | **2** |  | **4** |
| 7 | 数据可视化入门 | **2** | **2** |  | **4** |
| 8 | 城市大数据分析应用 | **2** |  |  | **2** |
| **合计** |  | **16** | **16** |  | **32** |

四、教学内容

第一章 大数据概念与Python基本语法

第一节 大数据概念

第二节 数据类型

第三节 变量

第四节 运算符

第五节 列表与元组

第六节 字符串

第七节 字典与集合

课程思政切入点：开展大数据行业行业规范教育、加强学生职业操守与法律意识。

第二章 条件语句、循环与函数

第一节 if语句

第二节 while 循环

第三节 for循环

第四节 函数

第三章 数据读取与存储

第一节数据读取（csv，txt）

第二节 数据存储

第三节 第三方模块引用

第四章 文本分析初步

第一节 文本字数统计

第二节 字频统计

第三节 关键词统计与提取

第四节 jieba分词入门、停用词、自定义词典

第五章 网络爬虫入门

第一节 网页构造和爬虫原理

第二节 网页的获取与阅读

第三节 简单爬虫实例

第四节 正则表达式与网页源代码解析

第五节 mysql基本操作与爬虫数据存储

第六节 异步加载网站解析

第七节 更多爬虫实例：政府工作报告、微博豆瓣等

课程思政切入点：在政府工作报告案例的讲解中，融入四个自信内容。

第六章 数据处理

第一节 Pandas模块基本应用：分列、合并、正则表达式、分组、函数调用等

第二节 中国知网数据清理

第三节 天气数据数据清理

课程思政切入点：在中国知网数据清理的案例中，加强学生学术道德教育。

第七章 数据可视化入门

第一节 Pyecharts：基本操作、词云、地图

第二节 地图绘图：数字普惠金融指数

第三节 matplotlib

第八章 城市大数据技术应用

第一节 大数据技术在城市管理中的应用

第二节 智慧城市大数据建设应用案例

课程思政切入点：通过城市大数据技术应用案例，开展行业规范教育、加强学生学术道德、职业操守与法律意识。

五、考核方式、成绩评定

平时成绩10%（点名+课堂表现）+作业20%+学期论文（70%）

六、主要参考书及其他内容

Python集成开发环境：安装Anaconda（<https://www.anaconda.com>），使用其中的Jupyter Notebook（Anaconda装好后直接启动后可使用，可将所有代码和执行结果存放于一个ipynb文件中，且能编辑和重新执行代码，还支持Markdown等文本编辑标记语言）

参考书目：

[1] 嵩天, 礼欣, 黄天羽.《Python语言程序设计基础(第2版)》. 北京：高等教育出版社，2017.2

[2] 韦斯麦金尼(Wes McKinney)著, 徐敬一译.《利用Python进行数据分析》. 北京：机械工业出版社, 2018.6

参考网站：

[1] Python主站：https://www.python.org

[2] Python中文文档：https://docs.python.org/zh-cn

[3] 相关Python库官网文档

Python Requests: https://www.python-requests.org

Python Beautiful Soup:

https://www.crummy.com/software/BeautifulSoup

Python Scrapy: https://scrapy.org

Python Pandas: https://pandas.pydata.org

Python Pyecharts: https://pyecharts.org

Python Matplotlib: https://matplotlib.org

执笔人：史晨辰 教研室主任：　　 系教学主任审核签名：