

信息可视化

【Information Visualization】

一、基本信息

课程代码：【2055019】

课程学分：【2】

面向专业：【数字媒体技术】

课程性质：【系级专业选修课】

开课院系：【信息技术学院数字媒体技术系】

使用教材：

教材【无】

参考书目【Python 数据分析与可视化】

先修课程：【无】

并修课程：【无】

二、课程简介

《Python数据分析与可视化》是计算机相关中的一门数据分析入门课程，该课程介绍了Python语言在数据分析方面的应用，按照数据分析的步骤，从数据预处理、分析、可视化等方面介绍了数据分析的方式，并通过两个完整的数据分析实例进行相关知识的学习。主要内容包括Python基础知识介绍、数据预处理、数据分析基础工具NumPy、处理结构化数据工具Pandas、数据分析与知识发现、scikit-learn实现数据分析、Matplotlib交互式图表绘制、实例：科比职业生涯进球分析、实例：世界杯等。

本课程的目的与任务是使学生通过本课程的学习，从数据分析的基础理论知识入手，按照数据分析的基本流程循序渐进的学习数据分析知识，并使用Python编程进行实战操作。通过两个完整数据分析实例的学习，帮助学生更好的掌握数据分析技能，做到理论与实践相结合，方法与应用相结合。本课程除要求学生掌握数据分析的基础知识和相关Python库使用，更重要的是要求学生掌握理论与实践结合的学习方式，为更深入地学习打下良好的基础。

三、选课建议

Python数据分析与可视化适合数字媒体技术专业学生学习，因侧重于python语言的使用并可用python语言对案例进行分析，适合有一定计算机基础和编程理解能力的学生学习，建议大三下学期开课。

四、课程与专业毕业要求的关联性

专业毕业要求	关联
L011: 能领会用户诉求、目标任务, 正确表达自己的观点, 具有专业文档的撰写能。	
L021: 能根据环境需要确定自己的学习目标, 并主动地通过搜集信息、分析信息、讨论、实践、质疑、创造等方法来实现学习目标。	●
L031: 工程素养: 掌握数学、自然科学知识, 具有工程意识, 能结合计算机、数字媒体技术相关专业知解决复杂工程问题	●
L032: 软件开发: 掌握主流设计技术、程序设计思维以及相关数据库技术, 具备建设可运行于多种终端网站的能力。	
L033: 系统运维: 系统地掌握计算机硬件、软件的基本理论、基本知识, 具备保障系统运行与维护基本技能。	
L034: 素材采集与处理: 掌握数字媒体的基本理论、主流数字媒体应用软件使用技术, 具备素材的采集、存储、处理以及传输的能力。	
L035: 数据分析概念与基本概念, 并掌握 python 相关基础知识。	
L036: 可使用 python 语言对数据进行基本的梳理和整理, 将数据进行可视化。	
L041: 遵守纪律、守信守责; 具有耐挫折、抗压力的能力。	●
L051: 能与团队保持良好关系, 积极参与其中, 保持对信息技术发展的好奇心和探索精神, 能够创新性解决问题。	●
L061: 能发掘信息的价值, 综合运用相关专业知识和技能, 解决实际问题。	●
L071: 愿意服务他人、服务企业、服务社会; 为人热忱, 富于爱心, 懂得感恩。	
L081: 具有基本外语表达沟通能力, 积极关注发达国家和地区信息技术发展新动向。	

五、课程目标/课程预期学习成果

序号	课程预期学习成果	课程目标 (细化的预期学习成果)	教与学方式	评价方式
1	LO212	能搜集、获取达到目标所需要的学习资源, 实施学习计划、反思学习计划、持续改进, 以达到学习目标	理论课	理论考试
2	LO311	运用高等数学等学科知识解决复杂工程问题	案例实践	阶段性作业
3	LO412	诚实守信: 为人诚实, 新手承诺, 尽职尽责。	师生交流	案例展示
4	LO512	有质疑精神, 能有逻辑的分析与批判。	课内师生知识分享	随堂表现
5	LO612	能够使用适合的工具来搜集信息, 并对信息加以分析、鉴别、判断与整合。	大作业合作情况考查	课堂表现

六、课程内容

1. 数据分析概述与基本概念 (实践课时: 2 学时)

了解数据分析的定义和历史背景; 了解数据分析的目的和任务; 了解数据分析与数据挖掘的关系; 了解机器学习与数据分析的关系; 了解数据分析的基本步骤, 即数据收集、

数据预处理、数据分析与知识发现和数据后处理，以及每个基本步骤的概念与内容；了解Python在数据分析领域所具有的优点。

重难点：数据分析的流程。

2. Python 基础知识（实践课时：4 学时）

了解Python的发展史；了解Python及Pandas、scikit-learn、Matplotlib的安装；掌握Python的基础知识及Python的一些特性，包括缩进在Python的重要性、Python包的使用方法、注释的使用方法、Python语法知识；了解Pandas、scikit-learn、Matplotlib等重要Python库；了解Jupyter交互式数据科学与科学计算开发环境。

重难点：python语言中的语法知识。

3. 数据预处理（实践课时：2 学时）

了解数据的集中趋势、离散程度、相关性测量、数据缺失、噪声、离群点等概念；了解数据质量标准评估的完整性、一致性、准确性和及时性4个要素。了解数据清洗的概念及方式，包括缺失值的处理、噪声数据的处理、不一致数据的处理和异常数据的处理；了解数据的特征选择、特征构建和特征提取。

重难点：数据预处理步骤，对异常数据的处理。

4. NumPy——数据分析基础工具（实践课时：4 学时）

了解NumPy库的作用；掌握多维数组对象ndarray的使用，包括ndarray的创建、ndarray的索引、切片和迭代、ndarray的shape的操作、ndarray的基础操作等。

重难点：numpy工具包的使用基本操作。

5. Pandas——处理结构化数据（实践课时：4 学时）

了解Pandas库的作用和特色功能；了解Pandas中Series和DataFrame两种主要数据结构；掌握Series对象的创建、访问和操作；掌握DataFrame对象的创建、访问和操作；掌握基于Pandas的Index对象的访问操作；了解Pandas的相关数学统计和计算工具；了解Pandas的数学聚合和分组运算。

重难点：pandas工具包的使用，进行简单的统计计算。

6. 数据分析与知识发现（实践课时：4 学时）

了解分类分析的基本概念；了解逻辑回归、线性判别分析、支持向量机、决策树、K近邻和朴素贝叶斯的相关知识；了解关联分析的基本概念，包括关联规则定义、指标定义和关联规则挖掘定义；了解关联分析的Apriori算法、FP-Tree算法、PrefixSpan算法；了解聚类分析的概念；了解K均值算法和DBSCAN；了解回归分析的基本概念和常用方法。

重难点：数据分析相关算法的了解。

7. scikit-learn——实现数据的分析（实践课时：4 学时）

了解SciPy和scikit-learn的作用；掌握scikit-learn的分类方法、回归方法和聚类方法，分类方法包括Logistic回归、SVM、最近邻分类器、决策树、随机梯度下降、高斯过程分类、神经网络分类和朴素贝叶斯，回归方法包括最小二乘法、岭回归、Lasso、贝叶斯岭回归、决策树回归、高斯过程回归和最近邻回归，聚类方法包括K-means、Affinity propagation、Mean-shift、Spectral clustering、Hierarchical clustering、DBSCAN和Birch。

重难点：数据分析算法的程序理解。

8. Matplotlib——交互式图表绘制（实践课时：4 学时）

了解Matplotlib的作用；掌握Matplotlib中的基本布局对象的建立；掌握修改图表样式、修改装饰项和添加注释的方法；掌握基础图表的绘制，包括直方图、散点图、柱状图、折

线图和表格；了解使用Matplot3D绘制三维图形的方法；了解Matplotlib与Jupyter结合使用的方法。

重难点：使用matplotlib工具包绘制图表。

9. 数据分析实例（实践课时：4学时）

通过“用户流失的预警”和“机器人走迷宫的最优路径”两个数据分析实例的分析，了解完整的数据分析实例的分析流程；了解实例中对数据的预处理、分析和绘图的过程；掌握综合运用所学知识，进行完整数据分析的能力。

重难点：独立分析数据的思维，熟悉数据分析流程。

七、课内实验名称及基本要求

序号	实验名称	主要内容	实验时数	实验类型	备注
1	Python 语言及其常用数据库	了解 python 语言的基础知识，包括数据处理工具包 numpy 和 pandas 等。	16	设计型	实验报告
2	数据可视化案例	了解数据可视化流程，完成可视化案例。	16	设计型	实验报告

八、评价方式与成绩

总评构成 (X)	评价方式	占比
X1	理论考试	40%
X2	阶段性作业（小组作业案例展示及实验报告）	30%
X3	平时成绩（考勤、课内练习及课堂表现）	30%

撰写人：徐红

系主任审核签名：张贝贝

审核时间：2023.2