《Python数据分析》教学大纲

**课程名称：**Python数据分析

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**64学时（其中理论36学时，实验28学时）

**总学分：**4.0学分

# 课程的性质

大数据时代已经到来，在商业、经济及其他领域中基于数据和分析去发现问题并做出科学、客观的决策越来越重要。数据分析技术将帮助企业用户在合理时间内获取、管理、处理以及整理海量数据，为企业经营决策提供积极的帮助。数据分析作为一门前沿技术，广泛应用于物联网、云计算、移动互联网等战略新兴产业。有实践经验的数据分析人才已经成为了各企业争夺的热门。为了推动我国大数据，云计算，人工智能行业的发展，满足日益增长的数据分析人才需求，特开设Python数据分析与应用课程。

# 课程的任务

通过本课程的学习，使学生学会使用Python进行科学计算、可视化绘图、数据处理，分析与建模，并详细拆解学习聚类、回归、分类三个企业案例，将理论与实践相结合，为将来从事数据分析挖掘研究、工作奠定基础。

# 课程学时分配

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 教学内容 | 理论学时 | 实验学时 | 其它 |
| 1 | 第1章 数据分析基础 | 1～2 |  |  |
| 2 | 第2章 NumPy模块实现数值计算 | 3～4 | 3 |  |
| 3 | 第3章 Pandas模块实现统计分析 | 3～4 | 4～5 |  |
| 4 | 第4章 Matplotlib模块实现数据可视化 | 2～3 | 3～4 |  |
| 5 | 第5章 【案例1】客户价值分析 | 2～3 | 4～5 |  |
| 6 | 第6章 | 2～3 | 4～5 |  |
| 7 | 第7章 【案例3】二手房数据分析预测系统。 | 2～3 | 5～6 |  |
| 8 | 第8章 【案例4】智能停车场运营分析系统。 | 2～3 | 6 |  |
| 9 | 第9章 【案例5】影视作品分析 | 1～2 | 4 |  |
|  | 第10章 【案例6】看店宝 | 2～3 | 7 |  |
|  | 总计 | 36 | 28 |  |

# 教学内容及学时安排

### 理论教学

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **章节名称** | | **主要内容** | **学时** |
| 1 | 数据分析基础 | | 主要介绍什么是数据分析、常用数据分析方法和工具、数据分析流程和Python数据分析常用模块 | 2 |
| 2 | NumPy模块实现数值计算 | | NumPy模块中的数组对象、数据类型对象、数组的基本操作以及常用的运算函数等 | 2 |
| 3 | Pandas模块实现统计分析 | | pandas的数据结构、文本数据以及数据库的读取或写入、常用的数据处理操作、数据的分组与聚合以及数据的预处理工作 | 2 |
| 4 | Matplotlib模块实现数据可视化 | | 通过Matplotlib模块实现可视化图形的绘制流程，以及绘制条形图、折线图、散点图等可视化图形 | 3 |
| 5 | 【案例1】客户价值分析 | | 通过RFM模型和k-means聚类算法实现客户分类和客户价值分析 | 3 |
| 6 | 【案例2】销售收入预测 | | 通过最小二乘法和线性回归模型linear\_model.LinearRegression实现销售收入分析与预测 | 6 |
| 7 | 【案例3】二手房数据分析预测系统 | | 通过sklearn模块中的线性回归等机器学习算法实现二手房数据分析预测系统 | 6 |
| 8 | 【案例4】智能停车场运营分析系统 | | 主要通过时间模块与pandas模块实现智能停车场运营数据的分析，再通过图表的方式实现数据的可视化 | 6 |
| 9 | 【案例5】影视作品分析 | | 主要通过Python的爬虫技术爬取影视作品的评论，然后通过pandas对数据进行处理，再通过图表的方式实现数据的可视化 | 6 |
|  | 【案例6】看店宝 | | 主要通过Python的爬虫技术获取京东商城商品的相关数据，然后通过数据的分析、比较、计算等方式实现京东商城商品的预警 | 6 |
| 36 | |

### 实验教学

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **实验项目名称** | **实验要求** | **学时** |
| 1 | Python数据分析环境搭建 | 1. 在Windows/Linux系统上安装Anaconda 2. 掌握Jupyter Notebook的常用功能 | 1 |
| 2 | NumPy数值计算基础 | 1. 创建NumPy数组对象ndarray 2. 查看ndarray的常用属性 3. 花式索引ndarray 4. 变换ndarray的形态 5. 创建NumPy矩阵并使用；使用常见ufunc 6. 使用NumPy读写文件 | 2 |
| 3 | Matplotlib数据可视化基础 | 1. 掌握pyplot的基本绘图语法 2. 设置pyplot的动态rc参数 3. 绘制散点图 4. 绘制折线图 5. 绘制直方图 6. 绘制饼图 7. 绘制箱线图 | 2 |
| 4 | pandas统计分析基础 | 1. 读写数据库数据 2. 读写文本文件 3. 读写Excel文件 4. 查看DataFrame的常用属性 5. 查改增删DataFrame数据 6. 描述分析DataFrame数据 7. 转换字符串时间为标准时间 8. 提取时间序列数据信息 9. 加减时间数据 10. 使用groupby方法拆分数据 11. 使用agg，apply，transform方法聚合数据 12. 制作透视表 13. 制作交叉表 | 4 |
| 5 | pandas数据预处理 | 1. 堆叠、主键、重叠合并数据 2. 检测与处理重复值，缺失值，异常值 3. 离差标准化、标准差标准化 4. 小数定标标准化数据；哑变量处理类别型数据 5. 离散化连续型数据 | 4 |
| 6 | sklearn模型构建 | 1. 加载datasets模块自带数据集 2. 划分数据集 3. 使用sklearn转换器进行数据预处理与降维 4. 构建与评价聚类模型 5. 构建与评价分类模型 6. 构建与评价回归模型 | 4 |
| 7 | 航空公司客户价值分析 | 1. 处理数据缺失值与异常值 2. 构建航空客户价值分析的关键特征 3. 标准化LRFMC 5个特征 4. 构建K-Means聚类模型 5. 评价K-Means聚类模型 | 4 |
| 8 | 财政收入预测分析 | 1. 分析财政收入数据特征的相关性 2. 使用Lasso回归选取财政收入预测的关键特征 3. 使用灰色预测和SVR构建财政收入预测模型 4. 评价SVR模型 | 3 |
| 9 | 家用热水器用户行为分析与事件识别 | 1. 删除冗余特征 2. 划分用水事件 3. 确定单次用水事件的时长阈值 4. 构建用水行为特征 5. 筛选候选洗浴事件 6. 构建BP神经网络模型 7. 评价BP神经网络模型 | 4 |
| 学 时 合 计 | | | 28 |

# 考核方式

突出学生解决实际问题的能力，加强过程性考核。课程考核的成绩构成 = 平时作业（10%）+ 课堂参与（20%）+ 期末考核（70%），期末考试建议采用开卷形式，试题应包括基本概念、绘图、分组聚合、数据合并、数据清洗、数据变换、模型构建等部分，题型可采用判断题、选择、简答、应用题等方式。

# 教材与参考资料

### 教材

王浩，袁琴，张明慧．Python数据分析案例实战[M]．ISBN：978-7-115-52084-5. 北京：人民邮电出版社．2020．