东北大学秦皇岛分校本科课程教学大纲

一、课程基本信息

<table>
<tr><td>开课单位</td><td colspan="5">悉尼智能科技学院</td><td>课程编号</td><td colspan="3">3100313036</td></tr>
<tr><td>中文名称</td><td colspan="8">数据分析编程</td></tr>
<tr><td>英文名称</td><td colspan="8">Programming for Data Analysis</td></tr>
<tr><td>总学时</td><td>48</td><td rowspan="2">理论学时</td><td rowspan="2">48</td><td rowspan="2">实验学时</td><td rowspan="2">0</td><td rowspan="2">上机学时</td><td rowspan="2">0</td><td rowspan="2">设计学时</td><td rowspan="2">0</td><td rowspan="2">课外学时</td><td rowspan="2">0</td></tr>
<tr><td>学分</td><td>3</td></tr>
<tr><td>课程类型</td><td colspan="11">☑ 理论课程 ☐ 理论(含实验)课程 ☐ 实践类课程</td></tr>
<tr><td>课程类别</td><td colspan="11">☐ 通识类 ☐ 学科基础类 ☑ 专业方向类 ☐ 实践类</td></tr>
<tr><td>适用专业</td><td colspan="11">应用统计学</td></tr>
<tr><td>先修课程</td><td colspan="11">概率论与随机变量 Probability & Random Variables
应用回归分析 Regression and Linear Models
Python 程序设计 Programming with Python</td></tr>
<tr><td>建议开课学期</td><td colspan="11">☐1-1 ☐1-2 ☐2-1 ☐2-2 ☑3-1 ☐3-2 ☐4-1 ☐4-2 ☐5-1 ☐5-2
☐*（代表每学期都开设）</td></tr>
<tr><td>考核方式</td><td colspan="4">☐ 考试 ☑考查</td><td>成绩分制</td><td colspan="5">☑百分制 ☐五级制 ☐二级制</td></tr>
<tr><td>使用教材</td><td colspan="11">1. Wes McKinney, “Python for Data Analysis. Agile Tools for Real World Data”, 2nd edition, O'Reilly Media, 2017.</td></tr>
<tr><td>参考书
及教学资源</td><td colspan="11">1．韦斯·麦金尼，《利用 Python 进行数据分析》，机械工业出版社，2023
2．https://www.py4e.com/book.php
3．https://jakevdp.github.io/PythonDataScienceHandbook/.</td></tr>
<tr><td>其他说明</td><td colspan="11">无</td></tr>
</table>

二、课程简介

中文描述

本课程为应用统计学本科专业的一门专业平台课，是一门综合性很强的课程，主要向学生介绍可用于收集、操作、建模和分析广泛数据集（包括缺失值的数据集）的数据结构和编程技术。本课程使用 Python 编程语言学习如何处理数字、字符串和更复杂的数据格式，并根据数据执行基本的数学建模或统计分析。通过课程的学习，学生能对多个领域的数据结构的共同特征进行清晰的理解，为其他专业课程的学习、创新创业活动、课程设计和毕业设计的开展，以及从事相关行业奠定坚实基础。

English Description

As a comprehensive subject, Programming for Data Analysis is a major subject for undergraduate students majoring in Applied Statistics. Based on the Python programming language, the content about this subject includes how to handle numbers, strings, and more complex data formats, and perform basic mathematical modeling or statistical analysis. And, after the learning of this subject, students can understand the data structures and gain the programming techniques to collect, manipulate, model, and analyze a broad range of datasets (including datasets with missing values). Also, students can get a clear understanding of the common characteristics of data structures in multiple fields, which can lay a solid foundation for the study of other subjects, innovation and entrepreneurship activities, subject design and graduation projects, as well as engaging in related industries.

三、课程目标

1、知识目标：

1-1：能熟练使用 Python 标准库，会查找与任何 Python 模块相关的信息，并使用这些信息编写代码。

1-2：能使用简单算法生成/创建 Python 代码。

1-3：能综合数据分析结果，并有效地传达结果。

1-4：识别并使用合适的在线资源来提高学习者的 Python 语言知识。

2、能力目标：

2-1：能使用 Python 进行探索性数据分析。

2-2：能够展示用于管理、操作、分析和可视化各种数据格式的基本 Python 编码技能。

2-3：能设计一个合适的编程工作流程来分析各种问题的数据，包括以前从未见过的问题。

3、德育/素养目标：

3-1：培养具有自主或团队合作能力，展示对需要应用数学和统计学的现实问题的专业和负责任的分析。

3-2：培养责任感与团队协作精神，以及职业道德与行为规范。

表 1 课程目标与毕业要求的对应关系

毕业要求	指标点 Indicator point	SLOs	关联 Relevancy (H/M/L)
2. 问题分析：能够借助应用统计学的基本原理、方法和手段，识别、表达、并通过文献研究分析复杂实际问题，以获得有效结论。	2.1 能够借助应用统计学的基本原理、方法和手段，分析、识别、表达本专业相关的复杂实际问题。	1-1 ~ 1-4 2-1 ~ 2-3	H
	2.2 能够借助应用统计学的基本原理、方法和手段，针对复杂实际问题设计针对性的方案，并综合运用文献、科学理论和技术手段予以解决。	1-1 ~ 1-4 2-1 ~ 2-3	H
3. 设计/开发解决方案：能够设计针对复杂实际问题的解决方案，设计满足特定需求的系统、单元或流程，并能够在设计环节中体现创新意识，考虑社会、健康、安全、法律、文化以及环境等因素。	3.2 能够对不同设计方案进行比较和优化，在工作各环节中具有创新意识和批判意识，善于发现、分析、系统表述和解决实际问题。	1-1 ~ 1-4 2-1 ~ 2-3	M
4. 研究：能够基于科学原理并采用科学方法对复杂实际问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4.2 能够结合本专业知识对实验数据进行分析与解释，设计并优化实验方案，并通过信息综合得到合理有效的结论。	1-1 ~ 1-4 2-1 ~ 2-3	M
	4.3 能够追踪国际前沿技术动态，掌握本专业涉及的重要技术指标以及达到指标所需的技术途径。	1-1 ~ 1-4 2-1 ~ 2-3 3-1 ~ 3-2	H
5. 使用现代工具：能够针对复杂实际问题，开发、选择与使用恰当的技术、资源、现代信息技术工具，包括对复杂实际问题的预测与模拟，并能够理解其局限性。	5.3 能够针对本专业相关复杂实际问题，选择与使用恰当的技术、资源、现代信息技术工具。	1-1 ~ 1-4 2-1 ~ 2-3	L

四、主要教学内容

（一）理论教学

第 1 章：Basic: Syntax, IPython and Jupyter Notebook（支撑课程目标 1-1, 1-2）

【1】知识点：

(1) Basic to Python.

(2) Basic Syntax of Python.

【2】学习目标：

理解：Basic to Python.

掌握：Basic Syntax of Python.

【3】重点内容：

(1) Basic to Python.

(2) Basic Syntax of Python.

【4】难点内容：

(1) Basic Syntax of Python.

第 2 章：Data structure, Functions, and Files of Python（支撑课程目标 1-1, 1-2, 3-1）

【1】知识点：

(1) Data structure and Sequence.

(2) Functions.

(3) Files and Operators.

【2】学习目标：

理解：Data structure and Sequence.

掌握：Functions; Files and Operators.

【3】重点内容：

(1) Data structure and Sequence.

(2) Functions.

(3) Files and Operators.

【4】难点内容：

(1) Data representations.

【5】思政案例：

Files are a form of data storage, and sometimes these documents need to be kept confidential, or users can only perform a few standardized operations on the files. In this case, file operations become very important. Therefore, in daily life, we need to have a sense of security and confidentiality, not only to protect our privacy or secrets, but also to keep national secrets confidential.

第 3 章：NumPy Basic: Array, Vector and Their Calculation （支撑课程目标 1-1, 1-2）

【1】知识点：

(1) The ndarray Object of NumPy.

(2) Generate Pseudorandom.

(3) The u-functions to Process ndarray.

(4) Ndarray-Oriented Programing.

(5) File I/O to ndarray

【2】学习目标：

掌握：The ndarray Object of NumPy; Generate Pseudorandom; The u-functions to Process ndarray; Ndarray-Oriented Programing; File I/O to ndarray.

【3】重点内容：

(1) The ndarray Object of NumPy.

(2) Generate Pseudorandom.

(3) The u-functions to Process ndarray.

(4) Ndarray-Oriented Programing.

(5) File I/O to ndarray

【4】难点内容：

(1) The u-functions to Process ndarray.

(2) Ndarray-Oriented Programing.

(3) File I/O to ndarray

第 4 章：Introduction to Pandas（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3）

【1】知识点：

(1) The Data Structure of Pandas.

(2) Basic Functions About Pandas.

(3) Summary And Calculation of Descriptive Statistics.

【2】学习目标：

理解：The Data Structure of Pandas.

掌握：Basic Functions About Pandas; Summary And Calculation of Descriptive Statistics.

【3】重点内容：

(1) The Data Structure of Pandas.

(2) Basic Functions About Pandas.

(3) Summary And Calculation of Descriptive Statistics.

【4】难点内容：

(1) Basic Functions About Pandas.

(2) Summary And Calculation of Descriptive Statistics.

第 5 章：Data File: Loading, Saving, Format（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3）

【1】知识点：

(1) File’s Format.

(2) Loading and Saving Text Files and Binary Files.

(3) Communicate with Database.

(4) Communicate with Web API.

【2】学习目标：

理解：File’s Format.

掌握：Loading and Saving Text Files and Binary Files; Communicate with Database; Communicate with Web API.

【3】重点内容：

(1) Loading and Saving Text Files and Binary Files.

(2) Communicate with Database.

(3) Communicate with Web API.

【4】难点内容：

(1) Loading and Saving Text Files and Binary Files.

(2) Communicate with Database.

(3) Communicate with Web API.

第 6 章：Data Cleaning and Preparation（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3）

【1】知识点：

(1) Processing Missing Data.

(2) Data Conversion.

(3) Extending Data Types.

(4) String and Its Processing.

(5) Classifying Data

【2】学习目标：

掌握：Processing Missing Data; Data Conversion; Extending Data Types; String and Its Processing; Classifying Data.

【3】重点内容：

(1) Processing Missing Data.

(2) Data Conversion.

(3) Extending Data Types.

(4) String and Its Processing.

(5) Classifying Data

【4】难点内容：

(1) Processing Missing Data.

(2) Data Conversion.

(3) Extending Data Types.

(4) String and Its Processing.

(5) Classifying Data

第 7 章：Data Organization: Connection, Union, And Reshaping（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3, 3-1, 3-2）

【1】知识点：

(1) Hierarchical indexing.

(2) Joining and Merging Datasets.

(3) Reshaping and Perspective.

【2】学习目标：

掌握：Hierarchical indexing; Joining and Merging Datasets; Reshaping and Perspective.

【3】重点内容：

(1) Hierarchical indexing.

(2) Joining and Merging Datasets.

(3) Reshaping and Perspective.

【4】难点内容：

(1) Hierarchical indexing.

(2) Joining and Merging Datasets.

(3) Reshaping and Perspective.

第 8 章：Data Visualization（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3）

【1】知识点：

(1) Introduction to Matplotlib API.

(2) Visualization using Pandas and Seaborn.

(3) Other Tools to Visualization.

【2】学习目标：

掌握：Introduction to Matplotlib API; Visualization using Pandas and Seaborn; Other Tools to Visualization.

【3】重点内容：

(1) Introduction to Matplotlib API.

(2) Visualization using Pandas and Seaborn.

(3) Other Tools to Visualization.

【4】难点内容：

(1) Visualization using Pandas and Seaborn.

(2) Other Tools to Visualization.

第 9 章：Data Aggregation and Grouping Operations（支撑课程目标 1-1, 1-2, 1-3, 1-4, 2-1, 2-2, 2-3, 3-1, 3-2）

【1】知识点：

(1) The Mechanism of Group By.

(2) Data Aggregation.

(3) Pivot Table and Crosstab.

【2】学习目标：

掌握：The Mechanism of Group By; Data Aggregation; Pivot Table and Crosstab.

【3】重点内容：

(1) The Mechanism of Group By.

(2) Data Aggregation.

(3) Pivot Table and Crosstab.

【4】难点内容：

(1) The Mechanism of Group By.

(2) Data Aggregation.

(3) Pivot Table and Crosstab.

（二）实验教学

None

五、教学安排与学时分配

This subject mainly focuses on classroom teaching, supplemented by on-site practical teaching. Classroom teaching revolves around the basic features, concepts, and practical operations of Python and Pandas. The subject revolves around the fundamental task of cultivating morality and nurturing talents, with the overall goal of cultivating socialist builders and successors with comprehensive development in morality, intelligence, physical fitness, aesthetics, and labor. With the general method of data processing as the main line, the subject fully utilizes offline classrooms, online teaching resources, and ideological and political cases to carry out teaching activities. It plays a role in imparting theoretical knowledge, practical operation, and value guidance, cultivates patriotism and the concept of building a strong country, solidifies basic knowledge, enhances practical skills and innovative spirit, and establishes a correct outlook on life, world, and values. The allocation of offline teaching hours is shown in Table 2.

表 2 学时分配表

讲课内容		学时				
		讲课	实验	上机	设计	课外
第 1 章	Basic: Syntax, IPython and Jupyter Notebook	4	0	0	0	0
第 2 章	Data structure, Functions, and Files of Python	6	0	0	0	0
第 3 章	NumPy Basic: Array, Vector and Their Calculation	6	0	0	0	0
第 4 章	Introduction to Pandas	4	0	0	0	0
第 5 章	Data File: Loading, Saving, Format	4	0	0	0	0
第 6 章	Data Cleaning and Preparation	4	0	0	0	0

第 7 章	Data Organization: Connection, Union, And Reshaping	6	0	0	0	0
第 8 章	Data Visualization	6	0	0	0	0
第 9 章	Data Aggregation and Grouping Operations	8	0	0	0	0
小计		48	0	0	0	0
总计	48					

六、教学方法

1. Teach basic principles, apply theory to practice, emphasize the basic characteristics of Python language Pandas package, the basic ideas, and methods of Python programming, and cultivate students' innovative abilities.

2. The teaching method combines multimedia courseware with blackboard writing, offline self-learning and computer operation, and online Q&A with traditional Q&A to improve classroom vitality and Q&A ability.

3. Provide theoretical questions and experimental projects in advance, encourage students to learn on their own, stimulate their learning enthusiasm, cultivate their self-learning ability, problem-solving ability, and teamwork ability.

七、成绩评定

1. The exam form for this subject is open examination and will be evaluated on a hundred marks system. Combine attendance, assignments, and final exams to comprehensively evaluate grades.

(1) Attendance (10%): Students are required to participate in all stages of this subject and are not allowed to be absent.

(2) Homework (40%): Students are required to complete online assignments on time.

(3) Final exam (50%): Open examination on Computer, 120 minutes.

2. The assessment qualification will be cancelled for those who lack more than one-third of the required assignments, or are absent for more than one-third of the planned class hours.

八、评价及持续改进

1. Syllabus

1、教学大纲

Check Period: 4 years, Revised Period: 4 years.

考核周期：4 年，修订周期：4 年。

Improvement Measures: the subject coordinator is responsible to organize all members in subject team for the syllabus discussion and modification, and the revised version shall be submitted to director of the major then the deputy dean (teaching affairs) for reviewing and approval, and finally be executed.

改进措施：课程负责人组织课程团队所有教师讨论后提出改进意见，经教研室主任审核后由教学院长批准。

2. Evaluation

2、成绩评定考核

Check Period: 1 years. Evaluation method can be improved based on the comprehensive analysis of students' final scores.

考核周期：1 学年，评价依据：学生最终成绩综合分析。

Improvement Measures: based on the distribution of students' final scores, editor(s) can make adjustment and optimization of subject content, and improve teaching methods.

改进措施：根据学生最终成绩的分布，进行相应的内容调整和优化，并改进教学方法。

撰写：	张建波
审核：	韩鹏
批准：	史闻博
日期：	2023.07.05