**《巨量数据处理分析》教学大纲**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程名称：**巨量数据处理分析 | | | | | | | **课程类别（必修/选修）：**必修 | | | | |
| **课程英文名称：**Big data of data processing and analysis | | | | | | | | | | | |
| **总学时/周学时/学分：**48/3/3 | | | | | | | **其中实验/实践学时：**24/0 | | | | |
| **先修课程：Python相关课程** | | | | | | | | | | | |
| **授课时间：**1-16周 周一 1-3节 | | | | | | | **授课地点：**线上 | | | | |
| **授课对象：**2018计算机科学与技术(跨境电商)1班 | | | | | | | | | | | |
| **开课学院：**粤台产业科技学院 | | | | | | | | | | | |
| **任课教师姓名/职称：詹家榜** /副教授 | | | | | | | | | | | |
| **答疑时间、地点与方式：**  1.每次课的课前、课间和课后，采用一对一的问答方式；  2.每次习题课，采用集中讲解方式 | | | | | | | | | | | |
| **课程考核方式：**开卷**（）**闭卷**（✔）**课程论文**（）**其它**（✔）** | | | | | | | | | | | |
| **使用教材：Python+Spark 2.0+Hadoop机器学习与大数据实战（清华大学出版社）**  **教学参考资料： Hadoop Spark 大数据巨量分析与机器学习整合开发实战 大数据和机器学习的基本概念**  **优学院授课网站:**  https://courseweb.ulearning.cn/ulearning/index.html#/course/announcement?courseId=67891 | | | | | | | | | | | |
| **课程概述**  **在各项领域中充斥着不同类型的巨量数据，如何将这些大数据经过一系列的流程萃取出潜在的重要资讯成为在资讯领域中一项重要的课题，本课程将教授处理巨量数据的基础观念，透过Python 、Weka等目前主流大数据处理分析工具及程式语言培养学生运用机器学习、资料探勘以及统计等观念，设计不同大数据分析模型的能力，并且使用于创造电商在商业的价值，让本课程学生能够有拥有处理大数据与商业实战最重要两个面向之基本能力。**  **课程目标**  **1.经过修习本课程后，使学生能灵活运用大数据分析相关软体**  **2.使学生更了解大数据分析与应用之相关知识，以培养学生之专业能力，使其满足学界与业界对于大数据人才的需求。**  **3. 学生能够有拥有处理大数据与商业实战最重要两个面向之基本能力。**  **课程内容和要求**  **这门学科的知识与技能要求分为知道、理解、掌握、学会四个层次。这四个层次的一般涵义表述如下：**   1. **知道 : 是指对这门学科和教学现象的认知。** 2. **理解 : 是指对这门学科涉及到的概念、原理、策略与技术的说明和解释，能提示所涉及到的教学现象演变过程的特征、形成原因以及教学要素之间的相互关系。** 3. **掌握 : 是指运用已理解的教学概念和原理说明、解释、类推同类教学事件和现象。** 4. **学会 : 是指能模仿或在教师指导下独立地完成某些教学知识和技能的操作任务，或能识别操作中的一般差错。**   **课程基本要求是**  **（1）掌握巨量资料分析的基本概念和理论。**  **（2）掌握巨量数据处理技术和分析算法以及算法复杂性。** | | | | | | | | | | | |
| **课程教学目标**  **一、知识目标：**   1. 通过**巨量资料分析**的学习，应使学生系统地掌握**巨量资料分析**的原理、基本架构及各种应用. 2. 了解演**巨量资料分析**的发展趋势及在工业、商业和科学技术方面的广泛应用。   **二、能力目标：**   1. 锻炼学生应用各种手段查阅文献资料、获取信息、拓展知识领域、继续学习并提高业务水平的能力。 2. 通过学习，使学生掌握演算法 分析方法与设计的能力。   **三、素质目标：**   1. 理解“定量”的意义，理解产生测量误差的因素，对实验严格要求，从操作、记录、分析等环节培养学生认真的态度、科学的精神. 2. 培养学生具有主动参与、积极进取、崇尚科学、探究科学的学习态度和思想意识； 3. 养成理论联系实际、科学严谨、认真细致、实事求是的科学态度和职业道德。 | | | | | | | | **本课程与学生核心能力培养之间的关联(授课对象为理工科专业学生的课程填写此栏）：**  通过本专业的学习，学生应具备如下核心能力：  ■核心能力1（交叉知识的运用能力）：具有运用数学、基础科学及计算机科学与技术相关知识的能力；  ■核心能力2（实验与数据解读能力）：具有计算机软件开发与数据搜寻分析解释的能力；  ■核心能力3（技术工具的应用能力）具有计算器软件工程技术应用、数据搜集分析应用跨境电商运营知识与技能、及大数据技术的专业所需的技术、技能和使用软硬件辅助工具的能力；  ■核心能力4（计科与大数据分析专业能力）：具有编程设计能力并能应用计算器与数据分析科技来辅助、及大数据技术分析，促进跨境电商运营的能力；  ■核心能力5（项目管理与团队合作能力）具有项目管理、有效沟通、领域整合与团队合作的能力；  ■核心能力6（解决复杂问题的能力）：具有运用计算机科学与技术理论及应用知识，整合计算机应用技术、数据分析应用及跨境电商运营专业，解决相关问题和进行研发或创新的能力；  ■核心能力7（持续学习与创新超越能力）：具有应对计算器科学与技术快速变迁的能力，培养自我持续学习的习惯与能力，了解所学专业技术对环境、社会及全球的影响，并在学习中敢于创新超越。  □核心能力8（专业伦理、社会责任与国际视野）：具有理解职业道德、工程专业伦理、认知社会责任、国际观以及开拓全球视野及尊重多元观点的能力。 | | | |
| **理论教学进程表** | | | | | | | | | | | |
| **周次** | **教学主题** | | **主讲教师** | **学时数** | **教学的重点、难点、课程思政融入点** | | | **教学模式**  **（线上/混合式/线下** | **教学方法** | | **作业安排** |
| 1 | 巨量资料简介 | | 詹家榜副教授 | 3 | 重点：巨量资料定义  难点：巨量资料之相关应用  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 2 | 数据思维与AI商业个案解析 | | 詹家榜副教授 | 3 | 重点：学习AI行销学原因与商业角度下的机器学习概论。  难点：机器学习与商业关系之连结  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 3 | AI行销学个案 | | 詹家榜副教授 | 3 | 重点：个案的探讨  难点：由个案的探讨找出关键因素。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 4 | Python商务实用基础（1） | | 詹家榜副教授 | 2 | 重点：介绍Python基本概念。  难点：基础教学与使用 。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 5 | Python商务实用基础（2） | | 詹家榜副教授 | 2 | 重点：Pandas DataFrame资料处理与转换  难点：熟悉套件的使用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 6 | Python商务实用基础（2） | | 詹家榜副教授 | 2 | 重点：Function方法操作-模组化商务城市的整合 。  难点：熟悉Function模组之使用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 7 | 期中总结、复习与期中考 | | 詹家榜副教授 | 3 | 期中总结、复习与期中考 | | | **线上** | **讲授** | |  |
| 8 | 商务实战概念-行销步骤与RFM客户行销模型 | | 詹家榜副教授 | 2 | 重点：RFM客户分类模型实战 。  难点：分析RFM相关步骤。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 9 | 商务实战概念-行销步骤与RFM客户行销模型 | | 詹家榜副教授 | 2 | 重点：资料清理、处理与转换、客户区、推荐。  难点：资料预处理与模型建立  课程思政融入点：  培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 10 | 商务分类模型概念介绍 | | 詹家榜副教授 | 2 | 重点：分类模型探讨。  难点：机器学习分类模型。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 11 | Python-罗吉斯回归（Logistic Regression） | | 詹家榜副教授 | 2 | 重点：Logistic Regression。  难点：Logistic Regression相关应用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 12 | Python-随机森林树（Random Forest）  Python-XGBoost | | 詹家榜副教授 | 2 | 重点： 。Random Forest  难点：Random Forest相关应用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 13 | 电商零售购物分析 | | 詹家榜副教授 | 2 | 重点：1.情境与资料变数介绍  2.资料型态与编码转换3.利润模型概念解析4.节流省钱分析清单与评估模型。  难点：知道怎么让程序运作仅是入门最重要的事利润模型解析。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 14. | 期末专题小组报告 | | 詹家榜副教授 | 3 | 重点：将本次课程所教以小组方式找寻数据集进行分析  难点：小组成员中互相分配任务相互学习。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 15 | 期末专题小组报告 | | 詹家榜副教授 | 3 | 重点：将本次课程所教以小组方式找寻数据集进行分析  难点：小组成员中互相分配任务相互学习。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| 16 | 期末专题小组报告 | | 詹家榜副教授 | 3 | 重点：将本次课程所教以小组方式找寻数据集进行分析  难点：小组成员中互相分配任务相互学习。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **线上** | **讲授** | | 教材课后习题  课程思政作业：要求学生每人阅读巨量数据处理分析有关的文章或书籍 |
| **合计：** | | | | 39 |  | | |  |  | |  |
| **实践教学进程表** | | | | | | | | | | | |
| **周次** | **实验项目名称** | **主讲教授** | | **学时** | **重点、难点、课程思政融入点** | | | **项目类型（验证/综合/设计）** | **教学**  **手段** | | |
| 4-6 | Python商务实用基础（1-3） | 詹家榜副教授 | | 3 | 重点：介绍1.Python基本概念。  2.Pandas DataFrame资料处理与转换3. Function方法操作-模组化商务城市的整合。  难点：基础教学与使用 。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 8-9 | 商务实战概念-行销步骤与RFM客户行销模型 | 詹家榜副教授 | | 2 | 重点：Pandas DataFrame资料处理与转换  难点：熟悉套件的使用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 10 | 商务分类模型概念介绍 | 詹家榜副教授 | | 1 | 重点：Function方法操作-模组化商务城市的整合 。  难点：熟悉Function模组之使用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 11 | Python-罗吉斯回归（Logistic Regression） | 詹家榜副教授 | | 1 | 重点：Logistic Regression。  难点：Logistic Regression相关应用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 12 | Python-随机森林树（Random Forest）  Python-XGBoost | 詹家榜副教授 | | 1 | 重点： 。Random Forest  难点：Random Forest相关应用。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 13 | 电商零售购物分析 | 詹家榜副教授 | | 1 | 重点：1.情境与资料变数介绍  2.资料型态与编码转换3.利润模型概念解析4.节流省钱分析清单与评估模型。  难点：知道怎么让程序运作仅是入门最重要的事利润模型解析。  课程思政融入点：培养实事求是的科学态度和职业道德。 | | | **综合** | 线上 | | |
| 合计： | | | | 9 |  | | |  |  | | |
| **考核方法及标准** | | | | | | | | | | | |
| **考核形式** | | | | | | **评价标准** | | | | **权重** | |
| 平时成绩 | | | | | | 出勤状况  1. 评价标准：不迟到，请假须有辅导员签字的请假条。  **线上教学**课堂表现  1. 评价标准：参与优学院学习参与成绩,课堂程度及随堂测验。  2. 要求：参与课堂程度高。  平时作业  1. 评价标准：按照作业完成情况评分。  2. 要求：按时作业，作业工整规范。 | | | | 40% | |
| 期中考试 | | | | | | 1. 评价标准：按照试卷参考解答及评分标准给分。  2. 要求：能灵活运用所学算法基础知识和方法进行求解，独立、按时完成考试。 | | | | 20% | |
| 期末报告 | | | | | | 1. 评价标准：按照试卷参考解答及评分标准给分。  2. 要求：能灵活运用所学算法基础知识和方法进行求解，独立、按时完成考试。 | | | | 40% | |
|  | | | | | |  | | | |  | |
|  | | | | | |  | | | |  | |
| **大纲编写时间：**2020**年**9**月8日** | | | | | | | | | | | |
| **系（部）审查意见：**  系（部）主任签名： 时维宁  日期：2020 年 9 月 7 日 | | | | | | | | | | | |