

数据科学与大数据技术微专业培养方案

一、专业名称

数据科学与大数据技术微专业

二、专业简介

数据科学与大数据技术微专业为非数据科学与大数据技术专业学生开设。面向互联网、大数据、云计算、人工智能等信息时代发展需要，培养掌握数据科学基本理论和数据分析处理技术，具备大数据技术应用开发与分析能力，能够利用大数据思维解决主专业大数据应用问题的复合型人才。

三、培养目标

培养德、智、体、美、劳等全面发展，具有扎实的计算机、数学、统计学等多学科的理论和专业基础知识，掌握数据科学与大数据技术的基本理论、基本方法和技能，具有创新能力、较强的工程实践能力和团队协作能力，能在多学科环境下从事大数据的采集与处理、存储与管理、分析挖掘、展现，和系统应用与运维等工作，适应现代化建设和未来社会科技发展需要的复合型应用人才。

四、培养要求

1. 掌握大数据分析与应用的基本理论和知识。
2. 具有计算思维和互联网思维能力，具有自主学习和终身学习意识，持续了解大数据分析技术的前沿和趋势；
3. 具有一定的工程创新能力和应用能力，具备较好的沟通、协调和合作能力和可持续发展能力；
4. 能够结合产业需求及市场实际，能够选择、开发与使用恰当的技术、资源和工具，对复杂大数据工程问题的预测、模拟和实现。

五、修读年限

基本修读年限为 1.5 年。主修专业毕业或结业，微专业学业自然终止。

六、结业标准

学生完成微专业培养方案规定的全部内容，成绩合格，达到微专业培养要求的，可获得微专业结业证书。

专业负责人：

年 月 日

分 管 院 长：

年 月 日

分 管 校 长：

年 月 日

数据科学与大数据技术微专业培养方案课程安排

课程代码	课程名称	开课学院	学分	学时	学时分配			开课学期	学分要求
					理论	上机	实验实践		
160527C040	Java 程序设计	石油学院	3	48	32	16		二	10
160527C057	数据统计与分析	石油学院	2	32	24	8		三	
160527C056	Hadoop 大数据应用	石油学院	3	48	32	16		六	
160527C030	机器学习	石油学院	3	48	40	8		六	
100514T024	数据可视化与应用	石油学院	2	32		32		五	4
160527C055	大数据分析算法（双语）	石油学院	3	48	32	16		六	
合计									14

课程简介：

Java 程序设计：Java 语言是目前功能最强、应用最广泛的一种完全面向对象程序设计语言，具有面向对象、平台无关、多线程、适合网络编程等特点，近年得到迅速推广应用。Java 语言功能丰富，表达能力强，使用灵活方便，应用广，目标程序可移植性强。本课程是计算机类工科各专业学习的一门高级语言程序设计课程，具有计算机语言基础即可学习本课程。通过本课程的学习，学生应掌握 Java 语言的基本语法，还应掌握面向对象程序设计的基本思想、基本概念和基本方法，并能运用所学的知识和技能对一般问题进行分析和程序设计，编制出高效的 Java 语言应用程序，为学习以后的课程如 Hadoop 大数据分析 & 大数据分析算法等打好基础。

数据统计与分析：设置本课程的目的在于培养学生有关统计方面的基本知识和技能，培养学生应用统计方法分析和解决问题的实际能力。教学应达到的总体目标是：

使学生能系统地掌握各种统计方法，并理解各种统计方法中所包含的统计思想。

使学生掌握各种统计方法的不同特点、应用条件及适用场合。

培养学生运用统计方法分析和解决实际问题的能力。

本课程在数据科学与大数据技术专业本科生的培养过程中，起着重要的基础作用。

Hadoop 大数据分析：本课程以“构建知识体系、阐明基本原理、引导初级实践、了解相关应用”为原则，为学生在大数据领域指明方向。课程将系统讲授大数据的基本概念、大数据处理架构 Hadoop、分布式文件系统 HDFS、分布式数据库 HBase、NoSQL 数据库、云数据库、分布式并行编程模型 MapReduce、数据仓库 Hive，大数据在各个领域的应用。

机器学习：本课程是人工智能技术的基础课程。本课程的教学目的是使学生理解机器学习的基本问题和基本算法，掌握它们的实践方法，为学生今后从事相关领域的研究工作或项目开发工作奠定坚实的基础。具体要求：理解聚类、回归、分类等相关算法并掌握它们的应用方法；理解神经网络类模型并掌握它们的应用方法；理解深度学习模型并掌握它们的应用

方法；理解距离度量、模型评价、过拟合、最优化等机器学习基础知识；掌握特征工程、降维与超参数调优等机器学习工程应用方法。通过编程练习和典型应用实例加深对算法的理解。

数据可视化与应用：数据可视化是关于数据视觉表现形式的科学技术研究。这种数据的视觉表现形式被定义为，一种以某种概要形式抽提出来的信息，包括相应信息单位的各种属性和变量。数据可视化是大数据内在价值的最终呈现手段,它利用各类图表、趋势图、视觉效果将巨大的、复杂的、枯燥的、潜逻辑的数据展现出来,使用户发现内在规律,进行深度挖掘,指导经营决策。

大数据分析方法：通过课程学习，使学生掌握大数据处理和应用中相关算法设计与分析的理论和方法。本课程的内容包括大数据分析的概念、大数据分析过程、大数据分析模型、数据挖掘与关联分析模型、分类分析模型、聚类分析模型、结构分析模型、文本分析模型、降维和数据仓库,通过理论讲授和作业分析相结合的方式培养学生对大数据分析方法及模型的兴趣，提高学生利用大数据分析模型解决现实问题的能力。