**大数据分析微专业课程介绍**

**1、回归分析**

回归分析主要研究如何从数据出发，揭示变量间的相互联系。它在经济、管理、工程、医药卫生等领域有着广泛的应用，也是大数据时代进行数据分析的基础工具。《应用回归分析》是统计学、大数据分析等专业的主干课程之一，课程主要内容包括古典的线性回归模型、回归诊断、非线性回归、Logistic回归等核心知识。在介绍回归分析的经典理论和实用方法的基础上，本课程面向时代发展，将传统的数理统计方法与机器学习的最新技术有机结合，并使用大数据时代最流行的编程语言——Python作为教学语言，使学生能够运用最新的分析工具进行数据挖掘。每个单元并配有适量的习题和实践案例，可供教师和学生使用。

**2、数据可视化**

数据可视化就是将各种类型的数据通过统计软件把它变为各类统计图。该课程重点讲解如何利用 Python 里面的绘图函数画出各类统计图，并且利用图形解释数据蕴含的统计规律。

**3、统计计算**

统计计算是现代统计的重要组成部分，它是统计方法和实际计算的结合。本课程以Python语言为基础，讲述基本的统计计算方法，包括：1、随机数抽样、参数估计和假设检验的数值计算；2、降低方差的统计方法，包括对偶变量、控制变量法、条件期望法、分层抽样法及随机数的重复应用等不同降低方差的思路及其应用。3、统计计算前沿方法，如Bootstrap、EM 算法、MCMC 方法等。使得学生初步掌握利用Python完成统计计算中算法的实现，把计算工具、统计方法、计算方法结合在一起解决数据分析问题。

**4、统计机器学习**

近年来统计机器学习方法已在各个行业得到广泛应用，反过来也极大地推动了新方法的大力发展。本课程通过大量实例系统介绍统计机器学习的核心方法及基础知识。内容包括：数据类型、数据预处理、概率统计模型、模型选择、数据拟合、监督学习、非监督学习、多重检验和贝叶斯统计。

**5、时间序列分析**

《应用时间序列分析》课程是应用统计学专业必修课，是该专业核心课程之一。该课程主要介绍了用R软件实现时间序列分析的主要方法和思想，是以概率论与数理统计为基础，在理解时间序列分析的基本原理的前提下，动手对时间序列数据进行分析和统计推断。时间序列分析在自然科学、管理科学和金融等领域应用十分广泛，通过课程的学习培养学生统计软件的操作能力和对时间序列分析方法的实际应用。本课程主要使学生掌握时间序列分析的基本理论和方法，侧重学生对分析方法的理解，从而使学生初步掌握分析时间序列数据的基本思路和方法，并能够用时间序列分析方法分析和解决、处理实际问题，为后续学习打下方法论基础。本课程融入课程思政教学，使学生深刻理解数据科学的发展历程，培养其树立正确的科学发展观，并坚定经世济国的使命和担当。

**6、金融大数据分析**

《金融大数据分析》课程将大数据分析、数据挖掘方法与金融市场学的相关知识相结合，以大数据分析及数据挖掘方法解决金融问题为导向，着重介绍大数据分析及数据挖掘方法在金融领域的应用。通过本课程的学习，使学生对金融市场知识及金融数据挖掘方法有比较系统和全面的了解；理清大数据分析、数据挖掘与金融的联系；理解应用大数据分析技术、数据挖掘等方法解决金融实际问题的思维方式；对大数据挖掘与金融相结合的前沿热点有正确的认识；培养学生具有开阔的眼界，以及思考和理解新概念的能力。

**7、商务大数据分析**

本课程从商务案例的数据分析出发，讲述常用统计方法、模型的应用。而且本课程还通过规范的商业分析报告展示数据分析流程和结果。本课程主要包括：线性回归——以移动通信网络的客户，方差分析——以北京市二手房价为例，逻辑回归——以上市企业特别处理 ST 为例，定序回归——以消费者偏好度研究为例，泊松回归——以付费搜索广告为例，生存数据回归——以员工离职管理为例等6章内容。通过本课程的学习，培养学生的基本商务数据分析思想，掌握商务数据统计分析概念基本原理，使学生初步掌握利用R软件完成统计模型的实现，具备运用统计方法和模型解决实际问题的能力。

**8、最优化方法**

最优化方法主要运用数学方法研究各种系统的优化途径及方案，为决策者提供科学决策的依据。主要研究对象是各种有组织系统的管理问题及其生产经营活动。研究目的在于针对所研究的系统，求得一个合理运用人力、物力和财力的最佳方案，发挥和提高系统的效能及效益，最终达到系统的最优目标。最优化方法被人们广泛地应用到公共管理、经济管理、工程建设、国防等各个领域，发挥着越来越重要的作用。

**9、非结构化数据统计分析**

《非结构化数据统计分析》是一门面向统计学、计算机、数据科学等相关专业的高级选修课程，主要涉及以下内容：①非结构化数据概述，介绍非结构化数据和结构化数据的区别，常见的非结构化数据类型以及获取途径；②自然语言处理基础，包括分词、词性标注和命名实体识别等自然语言处理技术；③文本挖掘与情感分析，介绍文本挖掘的概念和流程，以及如何进行情感分析。④图像与视频处理，主要介绍图像和视频处理的基本原理和技术。⑤音频信号处理，包括音频信号的获取和处理，以及语音识别、音乐分类等应用；⑥社交媒体数据分析，介绍社交媒体数据的获取和处理方法，以及如何进行网络分析和用户行为分析；⑦监督学习和无监督学习方法在非结构化数据分析中的应用。本课程旨在培养学生对非结构化数据的处理和分析能力，提升其大数据分析的实践能力。

**10、数据仓库与商业智能**

大数据时代亟需数据仓库与数据挖掘等技术集聚和挖掘数据资源，开发和释放数据蕴含的巨大价值，以数据竞争力支撑国家发展，以数据生产力推动社会进步。本课程将介绍数据仓库、联机分析处理（OLAP）工具和数据挖掘方面技术背景，并通过一系列实验综合运用这些技术。通过课程学习引导学生对数据挖掘产生兴趣，并且逐渐培养学生的数据意识，对较为成熟的技术和方法具有相应的分析应用能力。