附件一

2025级培养方案模板

数据科学与大数据技术专业专业学分制本科人才培养方案

（专业代码：080910T）

一、专业介绍

统计学院数据科学与大数据技术本科专业专注于统计与大数据分析交叉领域，采用跨专业、跨学科联合培养模式，旨在培育契合数据时代需求的综合型人才。

本专业发展历程丰富，2013年，统计学院联合多所高校筹建“大数据分析硕士培养协同创新平台”，并于次年招生，开创全国大数据硕士培养先河。同年，依托平台资源，开设“统计学（大数据分析）”本科方向，成为国内首批开展本科大数据分析教育的高校之一。2018 年，“数据科学与大数据技术” 本科专业获批招生。该专业在全国数据科学与大数据技术专业软科排名中连续获评 B+，并于 2020 年获批北京市一流本科专业建设点。

师资力量强劲，40 岁以下教师占比 63%，教授和副教授占比超 65%，科研方向涵盖数据分析各个领域。

专业培养特色显著。课程融合数学、统计学和计算机科学知识，培养学生跨学科思维，提升解决复杂数据问题的能力；重视实践教学，通过数据库原理与 SQL 实践等课程和项目实训，让学生在实操中掌握技能、解决实际问题；紧跟前沿技术，开设人工智能大语言模型等前沿课程，培养学生创新与竞争能力；与政府、金融、互联网企业紧密合作，借助实习、实践项目和企业讲座，让学生了解行业需求，提升职业素养和就业竞争力；依据学生兴趣特长，提供丰富选修课程与实践、科研机会，支持学生自主选择发展方向，实现个性化成长。

毕业去向多元，深造率超 50%，众多学生进入北京大学、新加坡国立大学等国内外顶尖院校。就业集中在互联网、金融、医疗、企事业单位等，主要从事数据分析、人工智能等工作，国有大中型企事业单位占比近 80%。

本专业社会声誉良好，为行业输送大量优秀人才，在大数据人才培养领域发挥重要引领作用。

二、培养目标

本专业致力于培养具有正确的社会主义核心价值观，适应社会经济发展和大数据时代需求，具备扎实的数学与统计学基础，熟练掌握数据科学与大数据技术的基本理论、方法和技能的高素质复合型人才。

具体目标如下：

目标1：熟悉大数据专业领域相关政策及法律、法规，能够在本专业领域实践活动中理解并遵守职业道德和职业规范。

目标2：能够在数据科学相关领域，运用大数据技术和分析方法，独立承担数据分析项目，从数据中提取有价值的信息，为企业的战略决策、产品优化、市场拓展等提供数据驱动的解决方案；

目标3：能够熟练运用机器学习、深度学习等人工智能技术，对复杂的大数据进行建模和预测，开发具有创新性的数据产品或服务，满足不同行业对数据智能应用的需求；

目标4：具备良好的团队协作和沟通能力，能够在跨学科团队中发挥核心作用，与不同专业背景的人员合作，共同完成复杂的数据项目，推动行业数字化转型和创新发展；

目标5：具有持续学习和自我提升的意识，能够跟踪数据科学与大数据技术的前沿发展，不断更新知识体系，适应快速变化的技术和市场环境，在职业生涯中实现持续发展。

三、毕业要求

**1．知识要求**

掌握数学分析、高等代数等数学基础知识，为数据科学的理论学习和实践应用提供坚实的数学支撑；深入理解数据科学的概率基础、统计基础等统计学知识，能够运用统计方法进行数据建模和分析；熟悉 Linux 系统、C 语言、Python 等编程语言，掌握数据库原理与 SQL 实践、数据结构与算法等编程和数据管理知识，具备数据处理和程序开发的能力；系统掌握机器学习与数据挖掘、神经网络与深度学习、人工智能大语言模型等数据科学核心知识，能够运用这些技术进行数据分析和预测；了解分布式计算、时间序列、应用回归分析、多元统计分析等相关知识，具备处理复杂数据问题的能力。

**2．能力要求**

具备数据采集、清洗、预处理和存储的能力，能够从多种数据源获取高质量的数据，并进行有效的管理和维护；熟练掌握数据分析和建模的方法和工具，能够运用统计分析、机器学习等技术对数据进行深入分析，构建数据模型，解决实际问题；具有数据可视化的能力，能够将复杂的数据以直观、易懂的方式呈现给用户，为决策提供支持；具备算法设计和编程实现的能力，能够根据实际问题设计有效的算法，并使用编程语言进行开发和优化；拥有团队协作和沟通能力，能够在团队项目中与他人合作，有效地表达自己的观点和想法，共同完成项目任务；具备自主学习和创新能力，能够不断学习新的知识和技术，在数据科学领域进行创新和实践。

**3．素质要求**

通过学生体质健康测试是毕业条件之一。具有良好的道德品质和职业操守，遵守行业规范和法律法规，在数据处理和分析过程中保护数据安全和隐私；具备较强的责任心和敬业精神，对工作认真负责，能够承担工作压力，在规定时间内高质量地完成任务；拥有积极的团队合作精神，能够与团队成员相互支持、协作，共同实现团队目标；具有创新意识和批判性思维，能够对现有技术和方法进行思考和改进，提出创新性的解决方案；具备良好的国际视野和跨文化交流能力，能够了解国际数据科学领域的最新发展动态，与国际同行进行交流和合作。

四、主干学科、核心课程及专业知识图谱

主干学科：统计学、数据科学

核心课程：数学分析、高等代数、数据科学的概率基础、数据科学的统计基础、Linux系统与c语言、数据库原理与SQL实践、数据结构与算法、机器学习与数据挖掘、数据准备、神经网络与深度学习、人工智能大语言模型、大数据分析实务、数据可视化、非结构数据分析与建模、分布式计算、时间序列、应用回归分析、Python数据分析、多元统计分析、统计编程基础、抽样技术与应用、最优化理论与方法、超高维数据分析、数值分析、统计预测与决策、应用随机过程、贝叶斯统计、应用非参数统计

专业知识图谱：

1. 核心知识领域

本专业方向的课程知识涵盖数学基础、计算机技术、数据科学核心技术三大方面。

①数学基础：数学分析、高等代数为数据科学提供了基础的分析工具和代数结构。数学分析中的极限、导数、积分等概念，是理解算法优化和数据变化趋势的基础；高等代数中的向量空间、矩阵运算等，在数据的表示与变换中应用广泛，如在图像处理、机器学习算法中的矩阵操作。数据科学的概率基础、统计基础是进行数据建模和分析的核心统计学知识。概率分布用于描述数据的不确定性，统计推断帮助从样本数据中得出关于总体的结论，为机器学习模型的训练和评估提供理论依据。

②计算机技术：Linux 系统与 C 语言、Python 数据分析等课程，让学生掌握不同编程语言在数据处理和程序开发中的应用。C 语言注重底层操作和算法实现，Python 则以其丰富的库和简洁的语法，成为数据处理和分析的主流语言。数据库原理与 SQL 实践、数据结构与算法是数据管理和处理的关键技术。数据库用于存储和管理数据，SQL 语言实现数据的查询和操作；数据结构和算法决定了数据的组织方式和处理效率，在大数据处理中发挥着重要作用。

③数据科学核心技术：机器学习与数据挖掘、神经网络与深度学习、人工智能大语言模型等，是实现数据分析、预测和智能应用的核心技术。机器学习算法用于从数据中自动学习模式和规律，数据挖掘则侧重于发现隐藏在数据中的知识；神经网络和深度学习在图像识别、自然语言处理等领域取得了巨大成功；大语言模型推动了自然语言处理的发展，为智能对话、文本生成等应用提供支持。大数据分析实务、数据可视化、非结构数据分析与建模、分布式计算等，针对大数据处理中的实际问题和应用场景。大数据分析实务涵盖从数据采集到结果呈现的全流程；数据可视化将复杂数据以直观图形展示，辅助决策；非结构数据分析与建模处理文本、图像等非结构化数据；分布式计算解决大数据处理的效率问题。

2、知识关联与层级结构

数学基础是整个知识体系的根基，为其他知识领域提供理论和方法支持。例如，机器学习算法中的梯度下降法基于数学分析中的导数概念，而数据在矩阵中的运算依赖高等代数知识 。计算机技术是实现数据处理和分析的工具和平台。通过编程语言和数据管理技术，将数学算法应用于实际数据处理，同时为数据科学核心技术的实现提供支撑。数据科学核心技术建立在数学和计算机技术基础之上，实现对数据的深度分析和智能应用，解决复杂的实际问题。各个核心技术之间也相互关联，如深度学习中会用到机器学习的一些基本概念和方法。

学生可依据知识图谱构建系统的知识体系，明确学习目标和路径。在学习过程中，理解知识间的关联，更好地掌握重点和难点知识，提升学习效果。教师借助知识图谱优化课程设置和教学内容安排，依据知识的层级结构和关联关系，合理安排课程顺序，使教学更具系统性和逻辑性。在科研和项目实践中，知识图谱帮助学生和研究人员快速定位所需知识，激发创新思维，促进知识的融合与应用，提升解决实际问题的能力。

**五**、学制和学位

学制：本专业基本学制为四年，实行弹性修业年限制度，学生在校修业年限可以提前至三年或延长至六年，修满规定的学分准予毕业。

学位：符合学士学位授予条件者，授予理学学士学位。

毕业条件：学生在规定的学习年限内，完成培养方案要求的最低总学分149学分。其中：

|  |  |  |  |
| --- | --- | --- | --- |
| **教学环节** | **课程类别** | **门数** | **学分** |
| 通识教育 | 通识教育必修课 | 22 | 49 |
| 通识教育选修课 |  | 10 |
| 专业教育 | 专业必修课 | 12 | 36 |
| 专业选修课 |  | 28 |
| 实践教育（不含课堂实验学分） | |  | 27 |
| 体质健康测试 | | 通过 | |
| **总计** | |  | 150 |

**六、学分一览表**

**总学分一览表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **总学分** | **教学方式** | | | | **理论与实践教学比例（%）** | |
| 149 | 课堂教学环节 | 123 | 理论教学 | 98 | 理论教学 | 66 |
| 实验教学 | 25 | 实践教学 | 34 |
| 课外教学环节 | 27 | 实习、军训等 | 27 |

1. **课堂教学学分分配表**

|  |  |  |  |
| --- | --- | --- | --- |
| **教学环节** | **课程类别** | **门数** | **学分** |
| 通识教育 | 通识教育必修课 | 22 | 49 |
| 通识教育选修课 |  | 10 |
| 专业教育 | 专业必修课 | 12 | 36 |
| 专业选修课 |  | 28 |
| **合计** |  |  | **123** |

1. **实践教学环节一览表**

|  |  |  |  |
| --- | --- | --- | --- |
| **类别** | **实践环节** | **学期安排** | **学分** |
| 实习类 | 军事技能 | 1 | 2 |
| 认知实习 | 4 | 2 |
| 专业实习 | 6 | 2 |
| 毕业实习 | 8 | 4 |
| 毕业设计（论文） | 8 | 4 |
| 素质提升类 | 创新学分 |  | 2 |
| 第二课堂 |  | 2 |
| 劳动类实践课程 |  | 2 |
| 人工智能导论 | 1 | 0.5 |
| 思想政治类 | 习近平新时代中国特色社会主义思想概论社会实践 | 1 | 1 |
| 思想道德与法治社会实践 | 2 | 1 |
| 毛泽东思想与中国特色社会主义理论体系概论社会实践 | 1 | 1 |
| 中国近现代史纲要社会实践 | 3 | 1 |
| 马克思主义基本原理社会实践 | 4 | 1 |
| 形势与政策（一） | 1 | 0.25 |
| 形势与政策（二） | 2 | 0.25 |
| 形势与政策（五） | 5 | 0.25 |
| 形势与政策（六） | 6 | 0.25 |
| 大学生心理健康实践 | 2 | 1 |
| 专业实验类 | Linux系统与c语言 | 3 | 2 |
| 数据科学的概率基础 | 2 | 1 |
| 数据科学的统计基础 | 3 | 1 |
| 数据库原理与SQL实践 | 5 | 1 |
| 时间序列分析 | 5 | 1 |
| 数据结构与算法 | 4 | 1 |
| 机器学习 | 5 | 1 |
| Python数据分析 | 3 | 2 |
| 分布式计算 | 6 | 1 |
| 大数据预处理 | 4 | 1 |
| 神经网络与深度学习 | 6 | 1 |
| 大数据分析实务 | 7 | 1 |
| 统计编程基础 | 3 | 2 |
| 多元统计分析（双语） | 5 | 1 |
| 超高维数据分析 | 7 | 1 |
| 最优化方法与应用 | 6 | 1 |
| 强化学习 | 6 | 1 |
| 非结构数据分析与建模 | 6 | 1 |
| 应用非参数统计 | 7 | 1 |
| 人工智能大语言模型 | 5 | 1 |
| 数值分析 | 6 | 1 |
| 可视化技术与实践 | 5 | 1 |
| **合计** | |  | **52.5** |