

《大数据导论》课程教学大纲

课程名称：大数据导论		课程编号：0632311	
课程类别：专业选修课		总学分：3	
总学时：54	理论学时：54		实践学时：0
适用对象：侦查学			
先修课程：概率论与数理统计、公安信息化技术			

一、教学目的及任务

本课程是侦查学的专业选修课，主要教学内容包括：大数据概述、大数据与其他新兴技术的关系、大数据基础知识、大数据应用、数据采集与预处理、数据存储与管理、数据处理与分析、数据可视化等内容，通过教学使学生能够了解大数据及相关技术的形成背景以及当今世界代表国家的大数据战略，理解并掌握大数据采集与预处理、数据存储与管理，数据处理与分析，数据可视化中的基本技巧、方法，相关技术架构与典型的平台，了解大数据思维和技术在互联网、物流、城市管理、金融、安全、政府等领域中的应用，形成运用大数据思维、模型和工具来解决实际问题的初步能力。

二、教学要求及内容

第一章 大数据概述

教学要求：通过本章内容的教学，使学生理解数据的概念、数据类型、数据组织形式、数据的价值；了解大数据的时代背景、发展历程、发展战略、影响、应用、产业和学科专业；理解大数据的概念定义与特性。

重点和难点：

1. 重点：人类 IT 发展史上 3 次信息化浪潮的发生时间、标志及其解决的问题；大数据的概念与特性。

2. 难点：大数据的特性。

教学内容：

第一节 数据

- 一、数据的概念
- 二、数据类型
- 三、数据组织形式
- 四、数据的使用
- 五、数据的价值
- 六、数据爆炸

第二节 大数据时代

- 一、第三次信息化浪潮
- 二、信息科技为大数据时代提供技术支撑
- 三、数据产生方式的变革促成大数据时代的来临

第三节 大数据的发展历程

- 一、大数据总体发展历程中的3个重要阶段

第四节 世界各国的大数据发展战略

- 一、美国、英国、法国、韩国、日本和中国的大数据发展战略

第五节 大数据的概念

- 一、数据量大
- 二、数据类型繁多
- 三、处理速度快
- 四、价值密度低

第六节 大数据的影响

- 一、大数据对科学研究的影响
- 二、大数据对社会发展的影响
- 三、大数据对就业市场的影响
- 四、大数据对人才培养的影响

第七节 大数据的应用

- 一、大数据在制造、金融、起床、互联网、餐饮、电信、能源等领域的应

用

二、从企业、政府、医疗等方面认识大数据的应用价值

第八节 大数据产业

一、大数据产业的不同层次及其包含内容

第九节 高校大数据专业

一、大数据专业的人才培养目标

二、毕业生就业岗位

三、大数据专业知识体系

四、大数据专业课程体系

五、大数据专业的编程语言

思考与练习：

1. 请从属性、比较、体系等角度出发，简述什么是大数据，其主要特征有哪些（即 5Vs）？
2. 请画图说明把数据变得可用需要经过哪个几个重要阶段？
3. 请简述大数据发展的 4 个重要阶段。

第二章 大数据与其他新兴技术的关系

教学要求：通过本章内容的教学，使学生了解什么是云计算、物联网、人工智能、区块链，它们包括哪些核心概念、关键技术、典型应用场景（产业），理解大数据与它们间、它们彼此间的关系。

重点和难点：

1. 重点：云计算、物联网、人工智能、区块链中的关键技术。
2. 难点：云计算、物联网、人工智能、区块链中的不同技术架构，以及需要解决的重点问题。

教学内容：

第一节 云计算

一、云计算的概念

二、云计算的服务模式和类型

三、云计算数据中心

四、云计算的应用

五、云计算产业

第二节 物联网

一、物联网的概念

二、物联网的关键技术

三、物联网的应用

四、物联网产业

第三节 大数据与云计算、物联网的关系

一、大数据与云计算、物联网之间的联系与区别

第四节 大数据与人工智能

一、人工智能的概念

二、人工智能的关键技术

三、人工智能的应用

四、人工智能产业

五、大数据与人工智能的关系

第五节 大数据与区块链

一、从比特币说起

二、区块链的原理

三、区块链的定义

四、区块链的应用

五、大数据与区块链的关系

思考与练习：

1. 请简述云计算有哪几种服务模式和哪几种类型？
2. 请画图说明物联网的技术架构，并阐述其中各个层次的具体功能。
3. 请阐述区块链是如何解决防篡改问题的。

第三章 大数据基础知识

教学要求：通过本章内容的教学，使学生理解大数据相关的一系列基础知识，包括大数据安全、大数据思维、大数据伦理、数据共享、数据开放、大数据交易中的相关概念，核心问题以及应用案例。

重点和难点：

1. 重点：大数据安全、大数据思维、大数据伦理、数据共享、数据开放、大数据交易中的重要任务、理论基础、面临的挑战和一些典型案例。

2. 难点：如何建立大数据思维方式；如何理解并缓解“数字鸿沟”、“数据独裁”、“数据垄断”问题。

教学内容：

第一节 大数据安全

- 一、传统数据安全
- 二、大数据安全与传统数据安全的不同
- 三、大数据安全问题
- 四、典型案例

第二节 大数据思维

- 一、传统的思维方式
- 二、大数据时代需要新的思维方式
- 三、大数据思维方式
- 四、运用大数据思维的具体实例

第三节 大数据伦理

- 一、大数据伦理的概念
- 二、大数据伦理典型案例
- 三、大数据的伦理问题

第四节 数据共享

- 一、数据孤岛问题
- 二、数据孤岛问题产生的原因
- 三、消除数据孤岛的重要意义
- 四、实现数据共享所面临的挑战
- 五、推进数据共享开放的举措
- 六、数据共享案例

第五节 数据开放

- 一、政府开放数据的理论基础
- 二、政府信息公开与政府数据开放的联系与区别

三、政府数据开放的重要意义

第六节 大数据交易

一、大数据交易概述

二、大数据交易发展现状

三、大数据交易平台

思考与练习：

1. 大数据安全与传统数据安全的不同之处体现在什么地方？
2. 请根据自己的生活实践列举一个大数据思维的典型案例。
3. 请阐述什么是“数字鸿沟”问题。
4. 请阐述什么是“数据独裁”问题。
5. 请阐述什么是“数据垄断”问题。
6. 简述数据质量的评估标准有哪些？
7. 请列举几个具有代表性的大数据交易平台。

第四章 大数据应用

教学要求：通过本章内容的教学，使学生了解大数据在各大领域中的典型应用。

重点和难点：

1. 重点：大数据在不同领域中的典型应用。
2. 难点：推荐系统中的长尾理论、推荐算法模型。

教学内容：

第一节 大数据在互联网领域的应用

一、什么是推荐系统

二、长尾理论

三、推荐方法

四、推荐系统模型

五、推荐系统的应用

第二节 大数据在生物医学领域的应用

一、流行病预测

二、智慧医疗

三、生物信息学

四、案例：基于大数据的综合健康服务平台

第三节 大数据在物流领域的应用

一、智能物流的概念

二、智能物流的作用

三、智能物流的应用

四、大数据是智能物流的关键

五、中国智能物流骨干网——菜鸟

第四节 大数据在城市管理领域的应用

一、智能交通

二、环保监测

三、城市规划

四、安防

五、疫情防控

第五节 大数据在金融领域的应用

一、高频交易

二、市场情绪分析

三、信贷风险分析

四、大数据征信

第六节 大数据在汽车领域的应用

一、无人驾驶汽车技术研究

第七节 大数据在零售领域的应用

一、发现关联购买行为

二、客户群体细分

三、供应链管理

第八节 大数据在餐饮领域的应用

一、餐饮领域拥抱大数据

二、餐饮 O2O

第九节 大数据在电信领域的应用

一、电信客户离网分析

第十节 大数据在能源领域的应用

一、大数据在智能电网、风力发电机安装选址方面的应用

第十一节 大数据在体育和娱乐领域的应用

一、训练球队

二、投拍影视作品

三、预测比赛结果

第十二节 大数据在安全领域的应用

一、大数据与国家安全

二、应用大数据技术防御网络攻击

三、警察应用大数据工具防范犯罪

第十三节 大数据在政府领域的应用

第十四节 大数据在日常生活中的应用

思考题

1. 请阐述什么是长尾理论。
2. 推荐系统的本质是建立用户与商品之间的联系，根据推荐算法的不同，请阐述推荐方法包括哪几类。
3. 请阐述智慧物流的概念和作用。
4. 请阐述大数据在金融领域有哪些典型应用。

第五章 数据采集与预处理

教学要求：通过本章内容的教学，使学生理解数据采集与处理的思路 and 流程，初步掌握大数据采集和预处理方法。

重点和难点

1. 重点：数据采集与预处理操作的要点与方法。
2. 难点：在真实场景下，如何针对具体问题选择并探索有效的数据采集与预处理方法。

教学内容：

第一节 数据采集

一、数据采集的概念

- 二、数据采集的三大要点
- 三、数据采集的数据源
- 四、数据采集方法
- 五、网络爬虫

第二节 数据清洗

- 一、数据清洗的内容
- 二、数据清洗的注意事项

第三节 数据转换

- 一、数据转换策略
- 二、平滑处理
- 三、规范化处理

第四节 数据脱敏

- 一、数据脱敏原则
- 二、数据脱敏方法

思考与练习：

1. 请阐述网络爬虫的组成。
2. 请阐述数据转换包括哪些策略？
3. 请阐述数据规范化包含哪些方法？
4. 请阐述数据脱敏的方法。

第六章 数据存储与管理

教学要求：通过本章内容的教学，使学生了解传统的数据存储与管理技术，在此之上理解大数据时代存储管理技术的变迁，掌握以分布式存储管理为核心的大数据存储与管理技术，同时了解其中的一些代表性产品。

重点和难点：

1. 重点：大数据时代的数据存储与管理技术。
2. 难点：NoSQL 数据库、云数据库和分布式数据库的数据模型与系统架构。

教学内容：

第一节 传统的数据存储与管理技术

- 一、文件系统

二、关系数据库

三、数据仓库

四、并行数据库

第二节 大数据时代的数据存储与管理技术

一、分布式文件系统

二、NewSQL 和 NoSQL 数据库

三、云数据库

第三节 大数据处理架构 Hadoop

一、Hadoop 特性

二、Hadoop 生态系统

第四节 分布式文件系统 HDFS

一、HDFS 的设计目标

二、HDFS 体系结构

第五节 NoSQL 数据库

一、键值数据库

二、列族数据库

三、文档数据库

四、图数据库

第六节 云数据库

一、云数据库的概念

二、云数据库的特性

三、云数据库与其他数据库的关系

四、代表性云数据库产品

第七节 分布式数据库 HBase

一、从 Bigtable 说起

二、HBase 简介

三、HBase 数据模型

四、HBase 系统架构

第八节 Google Spanner

一、Google Spanner 的特性与服务器的组织方式

思考与练习：

1. 请阐述键值数据库、列族数据库、文档数据库和图数据库的适用场合，以及各自的优缺点。
2. 请阐述在 Hadoop 体系架构中 HBase 与其他组成部分的相互关系。
3. 请阐述 Google Spanner 服务器的组织方式。

第七章 数据处理与分析

教学要求：通过本章内容的教学，使学生了解数据处理与分析的概念，理解典型机器学习和数据挖掘算法的原理与实施过程，掌握大数据处理与分析技术中的基本范式，以及代表性平台框架的设计思想、特点、工作流程，以及它们彼此间的关系。

重点和难点：

1. 重点：典型机器学习和数据挖掘算法和大数据处理与分析技术。
2. 难点：机器学习和数据挖掘算法的原理和大数据处理与分析框架的工作流程。

教学内容：

第一节 数据处理与分析的概念

- 一、数据分析与数据挖掘
- 二、数据分析与数据处理
- 三、大数据处理与分析

第二节 机器学习和数据挖掘算法

- 一、机器学习和数据挖掘算法概述
- 二、分类
- 三、聚类
- 四、回归分析
- 五、关联规则
- 六、协同过滤

第三节 大数据处理与分析技术

- 一、技术分类

二、流计算

三、图计算

第四节 大数据与分析代表性产品

一、分布式计算框架MapReduce

二、数据仓库Hive

三、数据仓库Impala

四、基于内存的分布式计算框架Spark

五、机器学习框架TensorFlowOnSpark

六、流计算框架Storm

七、流计算框架Flink

八、大数据编程框架Beam

九、查询分析系统Dremel

思考与练习：

1. 请简述 MapReduce 的工作流程。
2. 请简述 Spark 的体系架构包含哪些组件。
3. 为什么流式计算场景比较适合 Flink?
4. 简述 Beam 的设计目标。

第八章 数据可视化

教学要求：通过本章内容的教学，使学生了解数据可视化的类型和常用的可视化工具；掌握数据可视化的流程，处理和变换，练习形成数据可视化的评估能力。

重点和难点：

1. 重点：数据可视化的原理和过程，以及实现数据统计图表的实现。
2. 难点：数据可视化过程。

教学内容：

第一节 可视化概述

一、什么是数据可视化

二、可视化的发展历程

三、可视化的重要作用

第二节 可视化图表

一、最常用的统计图表类型及其应用场景

第三节 可视化工具

一、入门级工具

二、信息图表工具

三、地图工具

四、时间线工具

五、高级分析工具

第四节 可视化典型案例

一、全球黑客活动

二、互联网地图

三、编程语言之间的影响力关系图

四、世界国家健康与财富之间的关系

五、3D 可视化互联网地图 App

思考与练习：

1. 常见的统计图表有哪些类型？请给出每种类型的具体应用场景。
2. 可视化工具主要包含哪些类型？各自的代表性产品有哪些？
3. 请举出几个数据可视化的有趣案例。

第九章 大数据分析综合案例

教学要求：通过本章内容的教学，结合电影推荐系统具体案例，使学生对大数据分析全流程的部分理论和技术形成“串联”，掌握大数据理论和技术综合运用方法。

重点和难点：

1. 重点：电影推荐系统的系统设计方法。
2. 难点：电影推荐系统的算法设计。

教学内容：

第一节 案例任务

一、大数据实训案例——电影推荐系统

第二节 系统设计

- 一、系统总体设计
- 二、数据库设计
- 三、系统网站设计
- 四、算法设计

第三节 技术选择

- 一、电影推荐系统的实现技术
- 二、电影推荐系统的数据分析过程

第四节 系统实现

- 一、电影推荐系统的主要任务和实现

第五节 案例所需知识和技能

- 一、实现电影推荐系统案例所需知识和技能

思考与练习：

1. 简述大数据分析全流程主要包括哪些环节。
2. 简述最小交替二乘法（Alternating Least Squares, ASL）的基本原理。

三、教学方法与手段

（一）理论联系实际、课堂讲授与实际训练相结合的原则

为了使學生充分掌握大数据的基本理论，提高学生公安大数据专业素养，在教学中必须认真贯彻“理论联系实际”、“知识积累与技能发展并举”的教学原则，在深入浅出、生动形象、系统全面、科学准确讲授的同时，要通过案例教学和模拟教学等实训活动，给学生创造一个动脑、动口、动手的环境，使学生把学到的理论知识通过自己的实际演练，有效转化为实际工作的技能，从而达到为公安机关培养实用型专业人才的目的。

（二）精讲与自学相结合的原则

教学过程中应结合学生的实际情况，合理地确定教学重点和难点，精选讲授内容。对实践中必须掌握的基本知识、基本理论和基本技能要讲深讲透；对那些学生通过自学就能达到教学目的的内容，则应采取指导学生自学的方法，并借助于精选的教学参考资料，使学生全面地掌握课程的基本知识、基本理论和基本技能。

（三）教学形式多样、注重能力培养的原则

要紧密围绕当前公安业务工作实际需求，打破以教师、课堂、书本为中心的传统教学模式，根据教学内容和学生的实际情况，通过理论讲授、课堂讨论、平时作业、案例教学、模拟教学、多媒体教学等手段，帮助学生掌握本课程的基本理论、基本原理和基本技能。积极推广启发式、讨论式、探究式等教学方法，丰富教学训练载体，不断提高教学训练工作的针对性和实效性。充分发挥教师的主导作用和学生的主观能动性，努力做到教学内容与工作岗位、教学方法与工作方式、教学环境与工作氛围向一致，科学设计实训内容，强化实践性教学环节，培养和提高学生动手操作以及分析和解决实际问题的能力，以缩短适应实际工作岗位的时间。

（四）注重理论前瞻性，教书与育人相结合的原则

公安大数据涉及领域广，新情况、新问题不断出现，在教学过程中，要重视本专业领域的发展趋势，立足基层、贴近实战，把已经成熟或基本成熟的理论、方法和技术纳入教学内容，使学生了解本学科的最新发展和研究成果，为学生提供职业生涯发展的空间，努力培养学生参与社会实践的创新精神和职业能力。在教学过程中，教师要结合教学内容，挖掘德育渗透的因素，积极引导学生提升职业素养，提高职业道德水平。

（五）注重学习过程评价与学习结果评价相结合的原则

改革传统的学生评价手段和方法，注重学生学习过程和学习结果的全程评价。关注评价的多元性，在考试考核方式上，要把平时考核、实际训练和理论考试紧密结合，采用过程性评价与目标评价相结合、理论与实践一体化评价模式。注重学生动手能力和实践中分析问题、解决问题能力的考核，对在学习和应用上有创新的学生应予特别鼓励，全面综合评价学生的能力。

四、教学内容学时分配

序号	主要内容（写章目）	总学时	讲授学时	实践学时	备注
1	第一部分 大数据概述	6	6	0	
2	第二部分 大数据与其他新兴技术的关系	3	3	0	

3	第三部分 大数据基础知识	9	9	0	
4	第四部分 大数据应用	6	6	0	
5	第五部分 数据采集与预处理	6	6	0	
6	第六部分 数据存储与管理	6	6	0	
7	第七部分 数据处理与分析	9	9	0	
8	第八部分 数据可视化	6	6	0	
9	第九部分 大数据分析综合案例	3	3	0	
合计		54	54	0	

五、推荐教材及学习参考资源

（一）推荐教材

林子雨 编著. 《大数据导论》 中国工信出版集团 人民邮电出版社, 2020 年 9 月第一版

（二）学习参考资源

1. [英] 维克托·迈尔-舍恩伯格, [英] 肯尼思·库克耶 著, 盛杨燕, 周涛 译. 《大数据时代》 浙江人民出版社, 2013 年 1 月第一版
2. 车品觉 著. 《数据的本质》 北京联合出版公司, 2017 年 9 月第一版
3. 周涛 著. 《为数据而生: 大数据创新实践》 北京联合出版公司, 2016 年 4 月第一版
4. 王燃 著. 《大数据侦查》 清华大学出版社, 2018 年 7 月第一版
5. [土] 锡南·厄兹代米尔 迪夫娅·苏萨拉 著. 庄嘉盛 译 《特征工程入门与实践》 中国工信出版集团 人民邮电出版社, 2019 年 6 月第一版
6. [美] Michael Bowles 著. 沙赢 李鹏译 《Python 机器学习 预测分析核心算法》 中国工信出版集团 人民邮电出版社, 2017 年 1 月第一版
7. 李双其等著. 《大数据侦查实践》 知识产权出版社, 2019 年 9 月第 1 次印刷

（三）网络学习资源:

1. 厦门大学《大数据技术原理与应用》在线课程

<https://www.icourse163.org/course/XMU-1002335004>

2. 电子科技大学《大数据计算技术》

<https://www.icourse163.org/learn/UESTC-1003037002>

3. 微信公众号：“大数据分析 and 人工智能”、“大数据文摘”、“数据与算法之美”、“算法与数学之美”

六、考核方式

（一）考核类型：考试

（二）考核形式：笔试

（三）成绩评定方式：课程成绩=平时成绩（10%）+作业成绩（40%）+期末成绩（50%）。

课程负责人（签字）：

教研室主任（签字）：

系（部）主管领导（签字）：

系（部）（盖章）

年 月 日