

随着大数据时代的到来，隐私保护成为热门话题，得到社会各界的广泛关注。在数据科学项目中，需要注意保护用户隐私。隐私保护需要遵循相关的法律法规和伦理道德的要求。

7.9 继续学习本章知识

数据产品开发是学习和实践数据科学理论的主要抓手。本章主要介绍了数据产品的内涵与特征，继续学习本章知识应重视基于数据加工和数据柔术的数据产品开发的理论知识的学习和操作实践。数据思维是数据科学，尤其要注意数据产品开发与其他学科，如计算机科学中的软件产品开发的重要区别。因此，继续学习数据产品开发时也应注重培养自己的数据思维能力。

数据科学具有人文和管理属性，本章简要介绍了数据科学的项目管理、数据能力构建与评价、数据治理、数据安全、数据偏见、数据伦理和道德问题等。在后续研究中，需要进一步深入学习上述知识，并注重在实践中应用。在大数据时代，数据不仅是一种新的资源，更是一种重要资产，它涉及技术、经济、社会、法律、道德、伦理等多个领域。因此，我们应重视从跨学科视角学习和研究数据科学的理论知识与实践应用。

习 题

一、选择题

- () 指数据资源及其应用过程中相关管控活动、绩效和风险管理的集合。
A. 项目管理 B. 数据治理 C. 数据战略 D. 数据管理
- () 指组织中需要跨系统、跨部门进行共享的核心业务实体数据。
A. 核心数据 B. 业务数据 C. 主数据 D. 基础数据
- DMM 给出了组织机构数据管理所需的 25 个关键过程，并将其进一步聚类成 6 个关键过程域 ()、数据治理、数据质量、平台与架构、数据操作和辅助性过程。
A. 数据管理战略 B. 数据存储 C. 数据控制 D. 主数据管理
- DMM 中定义的关键过程域“数据质量”包括 ()。
A. 数据质量策略 B. 数据画像
C. 数据质量评估 D. 数据清洗
- 基于 DMM 模型的组织机构的数据管理能力成熟度水平的评价工作的实施可以借鉴 SEI 建议的 ()。
A. SET 模型 B. IDEAL 参考模型 C. SEI 模型 D. DIKW 模型



6. () 主要用于记录企业核心业务对象, 如客户、产品、地址等, 与交易流水信息不同, 这种数据一旦被记录到数据库, 需要经常对其进行维护, 从而确保其时效性和准确性。

- A. 主数据 B. 关系数据 C. 元数据 D. 交易数据

7. 数据治理任务通常包括 ()。

- A. 主动定义或序化规则
B. 为数据利益相关者提供持续跨界的保护、服务
C. 应对并解决因不遵守规则而产生的问题
D. 数据分析算法和模型的设计

8. () 指的是人们通常会注意到某种经过筛选之后所产生的结果, 同时忽略这个筛选的过程, 而被忽略的过程往往包含关键性的信息。

- A. A/B 测试 B. 大数据浮夸 C. 幸存者偏见 D. 大数据偏见

9. () 起源于 Web 测试, 是为 Web 或 App 界面或流程制作两个或多个版本, 在同一时间维度, 分别由属性或组成成分相同 (相似) 的两个或多个访客群组 (目标人群) 访问, 收集各群组的用户体验数据和业务数据, 最后分析、评估出最好版本, 将其正式采用。

- A. 版本测试 B. A/B 测试 C. 性能测试 D. 属性测试

10. () 是概率和统计学中的一种现象, 即几组不同的数据中均存在一种趋势, 但当这些数据组合在一起后, 这种趋势消失或反转。

- A. 辛普森悖论 B. A/B 测试 C. 幸存者偏见 D. 大数据涌现

11. 以下事件或行为中属于数据攻击的是 ()。

- A. Facebook 和剑桥分析公司的数据丑闻 B. 谷歌炸弹
C. 谷歌禽流感分析 D. 大数据杀熟

二、调研与分析题

1. 结合自己的专业领域或研究兴趣, 调研该领域的数据产品开发方法、技术与工具。

2. 分析 DMM 与 DAMA 数据管理知识体系 (DAMA Guide to the Data Management Body of Knowledge, DAMA-DMBOK) 的区别和联系。

3. 学习国家标准《数据管理能力成熟度评估模型》, 并分析与数据管理成熟度的异同。

