

# 数据科学应用

## 职业技能等级标准

(2022 年 1.0 版)

中国电子学会制定  
2022 年 3 月发布

# 目 次

前言.....	2
1 范围.....	3
2 规范性引用文件.....	3
3 术语和定义.....	3
4 适用院校专业.....	6
5 面向职业岗位（群）.....	7
6 职业技能要求.....	7
参考文献.....	13

# 前言

本标准按照GB/T 1.1-2020《标准化工作导则第1部分：标准化文件的结构和起草规则》的规定起草。

本标准起草单位：中国电子学会、深圳市深科数据科技有限公司、中国科学院计算机网络信息中心、中国（深圳）综合开发研究院、人民数据管理（北京）有限公司、北京易华录信息技术股份有限公司、凤凰数媒（北京）教育科技有限公司、深圳金证引擎科技有限公司、深圳瀚德金融科技控股有限公司、北京思源智通科技有限责任公司、上海和今信息科技有限公司等。

本标准主要起草人：许立达、孔晶、徐建、毛鹏清、沈志宏、李大卫、朱珍珠、巩冉、范向伟、孙正乾、李根柱、王阳、孙红民、吴晓琳、王宁桥、陈昕、刘国宏、胡彩梅、余凌曲、王娟、王海涛、王慧君、季婧、杨晋、吴艳光等。

**声明：本标准的知识产权归属于中国电子学会，未经同意，不得印刷、销售。**

## 1 范围

本标准规定了数据科学应用职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于数据科学应用职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

## 2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 19001-2016 《质量管理体系要求》

GB/T 19000-2016 《质量管理体系基础和术语》

GB/T 37973-2019 《信息安全技术 大数据安全管理指南》

GB/T 35294-2017 《信息技术 科学数据引用》

GB/T 35295-2017 《信息技术 大数据术语》

GB/T 34945-2017 《信息技术 数据溯源描述模型》

ISO/IEC 17000: 2004 《合格评定—词汇和通用原则》

ISO/IEC 20546:2019 《信息技术—大数据—概述和词汇》

ISO/IEC TR 20547-2018 《信息技术—大数据参考架构》

## 3 术语和定义

国家、行业标准界定的以及下列术语和定义适用于本文件。

### 3.1 控制流 Control Structures

控制流是计算机程序运行中决定程序运行方向和顺序的机制。

### 3.2 机器学习 Machine Learning

机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。

### 3.3 深度学习 Deep Learning

深度学习(DL, Deep Learning)是机器学习(ML, Machine Learning)领域中一个新的研究方向，它被引入机器学习使其更接近于最初的目标——人工智能(AI, Artificial Intelligence)。

### 3.4 自然语言处理 Nature Language Processing (NLP)

自然语言处理(Natural Language Processing, NLP)是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言，所以它与语言学的研究有着密切的联系，但又有重要的区别。自然语言处理并不是一般地研究自然语言，而在于研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统。因而它是计算机科学的一部分。

### 3.5 图像识别 Image Recognition

图像识别，是指利用计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术，是应用深度学习算法的一种实践应用。现阶段图像识别技术一般分为人脸识别与商品识别，人脸识别主要运用在安全检查、身份核验与移动支付中；商品识别主要运用在商品流通过程中，特别是无人货架、智能零售柜等无人零售领域。

### 3.6 人脸识别 Face Recognition

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术。用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸进行脸部的一系列相关处理技术，通常也叫做人像识别、面部识别。

### 3.7 无监督式学习 Unsupervised Learning

现实生活中常常会有这样的问题：缺乏足够的先验知识，因此难以人工标注类别或进行人工类别标注的成本太高。很自然地，我们希望计算机能代我们完成这些工作，或至少提供一些帮助。根据类别未知(没有被标记)的训练样本解决模式识别中的各种问题，称之为无监督学习。

### 3.8 监督式学习 Supervised Learning

监督式学习（英语：Supervised learning），是一个机器学习中的方法，可以由训练资料中学到或建立一个模式（函数 / learning model），并依此模式推测新的实例。训练资料是由输入物件（通常是向量）和预期输出所组成。函数的输出可以是一个连续的值（称为回归分析），或是预测一个分类标签（称作分类）。

### 3.9 强化学习 Reinforcement Learning

强化学习（Reinforcement Learning, RL），又称再励学习、评价学习或增强学习，是机器学习的范式和方法论之一，用于描述和解决智能体（agent）在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题。

### 3.10 能力 Competence

能力是经证实的能够应用知识和技能实现预期结果的特质。

### 3.11 职业技能 Professional Skills

职业技能指就业所需的技术和能力。

### 4 适用院校专业

院校	专业大类	专业代码	专业类
大专	电子与信息大类	510101	电子信息工程技术
		510202	计算机网络技术
		510207	信息安全技术应用
		510106	移动互联应用技术
		510203	软件技术
		510201	计算机应用技术
		510209	人工智能技术应用
		510206	云计算技术应用
		510205	大数据技术
		510208	虚拟现实技术应用
	财经商贸大类	530202	金融科技应用
530302		大数据与会计	
本科	计算机类	080910T	数据科学与大数据技术
		080901	计算机科学与技术
	电子信息类	080717T	人工智能
	数学类	070102	信息与计算科学
	统计学类	071201	统计学
		071202	应用统计学
	数学类	070101	数学与应用数学
	经济学类	020102	经济统计学
金融学类	020302	金融工程	
硕士	计算机类	0812J1	数据科学与大数据技术
		085211	大数据科学与工程

	理学类	70104	应用数学
		70105	运筹学与控制论

## 5 面向职业岗位（群）

主要面向科技、互联网、金融、电商、零售、交通、医疗、政务等领域，从事数据利用、数据挖掘、数据平台或系统端部署、应用，智能决策平台开发、部署、管理和优化，智能算法策划、开发、部署、管理和优化，数字化转型规划以及实施等相关岗位。

## 6 职业技能要求

### 6.1 职业技能等级划分

数据科学应用职业技能等级分为三个等级：初级、中级、高级，三个级别依次递进，高级别涵盖低级别职业技能要求。

**【数据科学应用】（初级）：**能够使用数据分析工具完成数据提取、数据清理、数据挖掘、数据分析、数据可视化、以及部署符合数据安全、数据道德规范的数字化软件。

**【数据科学应用】（中级）：**能够通过云平台操作、机器学习完成算法模型开发和不同应用场景需求的复杂数据项目实施、数据系统的开发与管理、部署与配置。

**【数据科学应用】（高级）：**能够通过深度学习完成高级模型开发和优化、不同应用场景需求的数据项目策略制定和规划、数据系统管理的优化、数据项目管理、指导、培训和评估。



## 6.2 职业技能等级要求描述

表 1 数据科学应用(初级)

工作领域	工作任务	职业技能要求
1. 数据项目可行性评估	1.1 业务调研	1.1.1 能根据业务流程，总结流程中重要的数据维度； 1.1.2 遵守数据规范，防止数据侵权违规； 1.1.3 能调研产业上下游并设定重要行业指标； 1.1.4 能使用编程语言和工具，（包括但不限于 Scrapy、Beautiful soup、Selenium）爬取行业公开数据。
	1.2 数据诊断	1.2.1 能根据数据项目的方向，判断所需重要数据维度； 1.2.2 能收集整理已有的数据集，从中判断数据项目所需维度是否满足； 1.2.3 能分析数据质量，包括数据缺失程度，数据源的准确性等； 1.2.4 能够对于多源异构的数据进行初步的转换。
	1.3 结论评估	1.3.1 能根据数据诊断结果评估数据项目实施可行性； 1.3.2 能对现有数据质量提出改善方案包括但不限于补充缺失，调整数据来源； 1.3.3 能编写数据项目所需的数据字典包括数据名称，数据简介，数据类型等； 1.3.4 能总结编写数据项目的可行性报告。
2. 数据项目开发	2.1 数据提取	2.1.1 能使用数据提取工具，包括但不限于 SQL、Python，从数据库中提取数据集； 2.1.2 能解析数据架构逻辑，决定数据提取方式； 2.1.3 能使用数据爬取能力，收集和整合非结构性数据； 2.1.4 能整合多个数据来源或系统，识别相关数据并组合多种类型数据。
	2.2 数据清理	2.2.1 能使用数据清理工具，包括：Linux、SQL、Python； 2.2.2 能用数据清理工具从数据集、数据表、数据库中识别数据的不完整、不正确、不准确或不相关部分，纠正被损坏或不准确的数据记录； 2.2.3 能用数据清理工具对数据进行转换和预处理（标准化、离散化、分层化）；

		2.2.4 能将不同类型的数据，如数字，文字，图片，视频类型进行转换成被模型兼容的格式。
	2.3 数据分析	2.3.1 能通过数据分析语言，包括但不限于 Python、R，以及其配套数据分析包实现基础模型搭建和应用； 2.3.2 能使用数据可视化工具，包括 Python、R、Tableau、PowerBI 等展现数据分析结果； 2.3.3 能通过数据分析语言，包括但不限于 Python、R，以及其配套数据分析包挖掘能助于业务决策的数据洞见； 2.3.4 能利用数据分析的洞见结合可视化图表，逻辑清晰的撰写支撑业务决策的报告。
3. 数据治理	3.1 数据安全	3.1.1 能在数据项目实施过程中遵守国内外数据安全相关法律法规； 3.1.2 能在数据项目开发中确保数据的可用性、完整性、保密性，保持数据安全意识； 3.1.3 保障数据处于有效保护和合法利用的状态； 3.1.4 保证数据处理的全过程安全，数据处理，包括数据的收集、存储、使用、加工、传输、提供、公开等。
	3.2 数据隐私	3.2.1 能遵守数据隐私国内外相关法律法规； 3.2.2 能在数据项目开发中做到个人隐私数据和公共隐私数据保护； 3.2.3 能对所涉及的隐私、敏感和机密数据分类； 3.2.4 能对“真实数据”以及“非翔实但真实的数据”或“被屏蔽数据”进行脱敏处理，并对访问敏感信息的人员实行持续监控和评估。
	3.3 数据伦理	3.3.1 能遵守数据伦理国内外相关法律法规； 3.3.2 能在数据项目开发中确保数据不滥采、不滥用； 3.3.3 能尊重自主原则，即数据的存储、删除、使用、知情等权利应充分理解并告知数据产生者。

表 2 数据科学应用(中级)

工作领域	工作任务	职业技能要求
1. 数据项目开发	1.1 云平台操作	1.1.1 能操作云平台，实现数据上云、算法上云； 1.1.2 能依据不同场景需求，根据各个云平台特征选择适配的平台；

		<p>1.1.3 能根据不同的场景需求，选择适配的云平台服务，如存储类型、实例规格、部署方案等；</p> <p>1.1.4 能利用云平台运行算法模型。</p>
	1.2 算法模型开发	<p>1.2.1 能运用 Python、R 等数据处理语言搭建算法模型；</p> <p>1.2.2 能操作数据科学计算机系统，包括关系型数据库操作、No-SQL 数据库操作、Spark 大数据处理；</p> <p>1.2.3 能开发统计模型机器学习，能用编程语言实现：线性回归和逻辑回归（二项式分布、多项式分布）、A/B 测试、方差分析（ANOVA）、时间序列（time series）等模型；</p> <p>1.2.4 能开发应用机器学习，用编程语言运行监督式学习（回归和分布式模型，支持向量机（SVM），决策树），无监督式学（聚类模型包括 K-Means 算法、DBSCAN 算法，降维技术包括主成分分析（PCA）、奇异值分解（SVD）、线性判别分析（LDA）、主题模型（topic modeling）），以及半监督式学习。</p>
	1.3 算法模型应用	<p>1.3.1 能针对应用场景，挑选符合场景需求的最优算法；</p> <p>1.3.2 能根据项目需求，拆分数据集：训练集、验证集和测试集，用于算法模型训练；</p> <p>1.3.3 能根据算法训练结果调整参数，挑选最适配的模型；</p> <p>1.3.4 能依据算法在应用场景中的表现，持续优化算法。</p>
2. 数据治理	2.1 数据系统开发与管理	<p>2.1.1 能依据数据项目在各种场景的要求在数据平台中配置对应的数据维度，能对项目所需的数据提取、清理、存储在数据系统内；</p> <p>2.1.2 能够结合不同业务场景数据的需求、法律法规，进行数据系统的功能升级与改善，包括但不限于：增加数据维度、增改数据标签类别；</p> <p>2.1.3 能设计数据系统开发流程，实现数据采集、整合、展现自动化；</p> <p>2.1.4 能将相关法律法规部署在数据项目规范、数据系统规则和数字化软件搭建等数据技术和管理措施中。</p>

	2.2 数据系统部署与配置	<p>2.2.1 能正确部署数据系统所需空间和配置；</p> <p>2.2.2 能制定数据系统部署策略，包括：数据维度、数据量、规则库等；</p> <p>2.2.3 能为所属数据项目适配数据系统功能选项。</p>
	2.3 数据安全管理与治理	<p>2.3.1 能遵守国家数据法规和规则；</p> <p>2.3.2 能采用数据备份、数据迁移、数据库加密等技术保障数据安全；</p> <p>2.3.3 能遵守企业数据安全制度。</p>
3. 数据项目管理	3.1 项目管理	<p>3.1.1 能总结项目结论报告；</p> <p>3.1.2 能根据项目需求，合理安排人员和其他资源；</p> <p>3.1.3 能根据数据日志（data log）把控项目质量，以及持续监控和升级。</p>
	3.2 培训与指导	<p>3.2.1 能将正确的数据科学基础理论与实务，对初级人员进行培训；</p> <p>3.2.2 能运用不同应用场景与案例，编写数据科学项目培训计划和实施技能测试；</p> <p>3.2.3 能依据数据项目报告，指导初级人员进行项目优化；</p> <p>3.2.4 能指导培训初级人员对于数据安全的理解和执行。</p>

表 3 数据科学应用(高级)

工作领域	工作任务	职业技能要求
1. 数据项目开发	1.1 算法模型开发和优化	<p>1.1.1 能运用 Python、R 等数据处理语言优化复杂算法模型，包括但不限于：深度学习如神经网络、激活和损失函数、正则化等；</p> <p>1.1.2 能将深度学习技术实现在包括：人脸识别、图像识别、自然语言处理、语音识别、推荐系统、决策智能化等业务需求中；</p> <p>1.1.3 能依据深度学习算法在应用场景中的表现，持续优化深度学习算法。</p>
	1.2 数据架构建设	<p>1.2.1 根据现阶段应用场景及未来发展方向构建数据框架；</p> <p>1.2.2 依据数据框架和业务流程，确定数据维度和类型；</p> <p>1.2.3 根据现有及未来数据量及数据时效性，确定存储方式（SQL/NoSQL）、储存地点（本地/云）等；</p>

		1.2.4 能结合业务，持续升级数据架构和开发新工具。
2. 数据治理	2.1 数据系统优化	2.1.1 能依托数据科学项目的方案和目标需求进行数据系统空间和功能优化； 2.1.2 能依据多部门、多数据源的需求，严格规范数据采集、清理、存储的流程，确保全流程可溯源、存储数据安全可用。
	2.2 数据安全	2.2.1 能把控评估数据系统（数据）的安全性和规范性； 2.2.2 能够迅速发现、反应、解决数据的安全隐患和安全问题； 2.2.3 能够对于数据安全的结果进行分析、审核、把控； 2.2.4 能应对数据安全的新法律法规提出解决方案。
3. 数据项目管理	3.1 项目部署	3.1.1 能设计数据项目策略与人员组； 3.1.2 能制定数据项目流程，并根据验证、检测结果及时调整流程规划； 3.1.3 能持续监督数据项目的运行，并优化项目结果为导向，持续升级技术、软件、人员的部署。
	3.2 项目管理	3.2.1 能应对项目实施过程中的突发状况，提供解决方案； 3.2.2 能总结数据项目实施经验，为建立流程化体系提供基础。
	3.3 培训与指导	3.3.1 能对初级、中级人员进行理论知识和操作技能的培训； 3.3.2 能利用数据科学项目的案例与经验，编写数据科学项目培训计划、技术文件和培训课件； 3.3.3 能运用不同特色应用场景，指导初级、中级人员进行模型优化、算法优化； 3.2.4 能指导培训中级人员对于数据安全的理解和执行。
	3.4 项目总结	3.4.1 能与业务需求部门或其他非数据部门有效沟通项目成果和项目价值，确保数据项目得到应用和落实； 3.4.2 能利用数据项目报告，总结算法模型优化方案； 3.4.3 能利用数据项目报告，进行数据科学从业人员的筛选，培训内容设计；

		3.4.4 能利用数据项目报告，设计数据工具和系统的升级改进方案。
--	--	-----------------------------------

## 参考文献

- [1]GB/T 19001-2016 质量管理体系要求
- [2]GB/T 19000-2016 质量管理体系基础和术语
- [3]GB/T 37973-2019 信息安全技术 大数据安全管理指南
- [4]GB/T 35294-2017 信息技术 科学数据引用
- [5]GB/T 35295-2017 信息技术 大数据术语
- [6]GB/T 34945-2017 信息技术 数据溯源描述模型
- [7]ISO/IEC 17000: 2004 合格评定—词汇和通用原则
- [8]ISO/IEC 20546:2019 信息技术—大数据—概述和词汇
- [9]ISO/IEC TR 20547-2018 信息技术—大数据参考架构
- [10]中等职业学校专业目录（2010年修订）
- [11]普通高等学校高等职业教育（专科）专业目录（截至2019年）
- [12]普通高等学校本科专业目录（2020年版）
- [13]中等职业学校专业教学标准
- [14]高等职业学校专业教学标准