

大数据应用开发（Python）

职业技能等级标准

标准代码：510050

（2021年2.0版）

广东泰迪智能科技股份有限公司 制定

2021年12月 发布

目 次

前言.....	1
1 范围.....	2
2 规范性引用文件.....	2
3 术语和定义.....	2
4 适用院校专业.....	3
5 面向职业岗位（群）	4
6 职业技能要求.....	4
参考文献.....	12

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：广东泰迪科技股份有限公司牵头发起，得到华为技术有限公司、湖北省工业与应用数学学会、朗新科技股份有限公司、网宿科技股份有限公司、广州粤嵌通信科技股份有限公司、蓝盾信息安全技术股份有限公司、深圳职业技术学院、广州番禺职业技术学院、深圳信息职业技术学院、武汉职业技术学院、江苏海事职业技术学院、河南工业职业技术学院、广东轻工职业技术学院、上海电子信息职业技术学院、浙江商业职业技术学院、大连职业技术学院、西安航空职业技术学院、广东科学技术职业学院、广东省人才研究会、北京四合天地科技有限公司等数十家大数据技术企业、行业学会、院校单位及专家学者的广泛参与支持。

本标准主要起草人：郝志峰、张良均、余明辉、詹增荣、张治斌、刘彦姝、秦宗槐、王津、王海、武春岭、施兴、赵云龙、蔡铁、陈永、杜恒、韩宝国、胡国胜、蒙飏、余爱民、史小英、沈凤池、沈洋、苏晓、万国德、张敏。

声明：本标准的知识产权归属于广东泰迪智能科技股份有限公司，未经广东泰迪智能科技股份有限公司同意，不得印刷、销售。

1 范围

本标准规定了大数据应用开发（Python）职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于大数据应用开发（Python）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 37696-2019 信息技术服务 从业人员能力评价要求

GB/T 36348-2018 信息技术 学习、教育和培训 虚拟实验 框架

GB/T 35295-2017 信息技术 大数据 术语

GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库

GB/T 5271.1-2000 信息技术 词汇 第1部分：基本术语

GB/T 5271.13-2008 信息技术 词汇 第13部分：计算机图形

GB/T 5271.31-2006 信息技术 词汇 第31部分：人工智能 机器学习

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标准。

3.1

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[GB/T5271.1-2000, 定义01.01.02]

3.2

大数据 big data

具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

[GB/T 35295-2017, 定义2.1.1]

3.3

关系数据库 relational database

数据按关系模型来组织的数据库。

[GB/T5271.17-2010, 定义17.04.05]

3.4

机器学习 machine learning

功能单位通过获取新知识或技能，或通过整理已有的知识或技能来改进其性能的过程。

[GB/T5271.31-2006, 定义31.01.02]

3.5

数据处理 data processing

数据操作的系统执行。

[GB/T5271.1-2000, 定义01.01.06]

3.6

数据管理 data management

在数据处理系统中，提供对数据的访问、执行或监视数据的存储，以及控制输入输出操作等功能。

[GB/T5271.1-2000, 定义01.08.02]

3.7

分析 analytics

根据信息合成知识的过程。

[GB/T 35295-2017, 定义2.1.48]

3.8

数据挖掘 data mining

从大量的数据中通过算法搜索隐藏于其中信息的过程。

[GB/T 33745-2017, 定义2.5.3]

3.9

可视化（用于计算机图形） visualization (in computer graphics)

为帮助人们理解,采用计算机图形和图像处理技术来表现各个过程或对象的模型或特性的做法。

[GB/T 5271.13-2008, 定义13.01.07]

3.10

操作系统 operating system

控制程序执行的软件,它能提供诸如资源分配、目录调度、输入输出控制及数据管理的服务。

[GB/T5271.1-2000, 定义01.04.08]

3.11

算法 algorithm

为解决问题严格定义的有限的有序规则集。

[GB/T5271.1-2000, 定义01.05.05]

3.12

深度学习 deep learning

深度学习是机器学习的分支,是一种以人工神经网络为架构,对数据进行表征学习的算法。

3.13

自然语言 natural language

一种其规则是基于当前的用法且无需特别规定的语言。

[GB/T5271.1-2000, 定义01.05.08]

4 适用院校专业

4.1 参照原版专业目录

中等职业学校:软件与信息服务、计算机应用、电子商务、网络营销、连锁经营与管理、客户信息服务、物流服务与管理、冷链物流服务与管理、会计、会计电算化、统计事务。

高等职业学校:大数据技术与应用、计算机信息管理、商务数据分析与应用、信息统计与分析、人工智能技术服务、财政、税务、财务管理、会计、审计、环境信息技术、智能产品开发、智能终端技术与应用、计算机应用技术、软件技术、软件信息与服务、电子商务技术、会计信息管理、统计与会计核算、商务管理、市场营销、电子商务、网络营销、物流管理、旅游管理、酒店管理、网络舆情监测。

高等职业教育本科学校:电子信息工程、计算机应用工程、软件工程、大数据技术与应用、金融管理、财务管理、会计、市场营销、电子商务、跨境电子商务、物流管理、旅游管理。

应用型本科学校：数据科学与大数据技术、大数据管理及应用、信息与计算科学、智能科学与技术、人工智能、数据计算及应用、计算机科学与技术、软件工程、电子商务、统计学、应用统计学、数学与应用数学、市场营销、数据警务技术、财务管理、物流管理、供应链管理、旅游管理。

4.2 参照新版职业教育专业目录

中等职业学校：大数据技术应用、软件与信息服务、计算机应用、电子商务、网络营销、连锁经营与管理、客户信息服务、物流服务与管理、冷链物流服务与管理、会计事务、统计事务。

高等职业学校：大数据技术、商务数据分析与应用、统计与大数据分析、人工智能技术应用、健康大数据管理与服务、财税大数据应用、大数据与财务管理、大数据与会计、大数据与审计、生态环境大数据技术、智能产品开发与应用、计算机应用技术、软件技术、会计信息管理、统计与会计核算、市场调查与统计分析、商务管理、市场营销、电子商务、网络营销与直播电商、现代物流管理、供应链运营、旅游管理、酒店管理与数字化运营、运动数据分析、网络舆情监测。

高等职业教育本科学校：电子信息工程技术、计算机应用工程、软件工程技术、大数据工程技术、人工智能工程技术、财税大数据应用、金融管理、大数据与财务管理、大数据与会计、大数据与审计、企业数字化管理、市场营销、电子商务、跨境电子商务、全媒体电商运营、现代物流管理、旅游管理。

应用型本科学校：数据科学与大数据技术、大数据管理及应用、信息与计算科学、智能科学与技术、人工智能、数据计算及应用、计算机科学与技术、软件工程、电子商务、统计学、应用统计学、数学与应用数学、市场营销、数据警务技术、财务管理、物流管理、供应链管理、旅游管理。

5 面向职业岗位（群）

【大数据应用开发（Python）】（初级）：主要面向各行业的数据管理、数据处理与分析、报表管理等领域的数据库专员、数据分析助理、数据专员、Excel 数据分析师、商业数据分析师等职业岗位，主要完成数据获取、数据处理、数据存储、数据分析、数据可视化、报表制作等工作。

【大数据应用开发（Python）】（中级）：主要面向各行业的平台管理与运维、大数据采集与存储、大数据分析与挖掘等领域的大数据运维工程师、大数据分析工程师、大数据挖掘工程师、数据可视化工程师、爬虫工程师等职业岗位，主要完成平台搭建与运维、大数据采集、大数据处理与存储、大数据分析挖掘、大数据可视化等工作。

【大数据应用开发（Python）】（高级）：主要面向各行业的平台高级配置及管理、大数据分析与挖掘、深度学习等领域的高级大数据分析工程师、高级大数据挖掘工程师、算法工程师、大数据架构师、人工智能工程师等职业岗位，主要完成大数据平台高级配置管理、大数据处理、大数据分析挖掘、模型优化与部署、文本挖掘、图像识别、语音识别等工作。

6 职业技能要求

6.1 职业技能等级划分

大数据应用开发（Python）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，工作领域和工作任务各有侧重。

【大数据应用开发（Python）】（初级）：根据业务需求和 workflows，主要使用 Excel、Power BI 和 Python 等工具，完成数据获取、数据处理、数据存储、数据分析、数据可视化、报表制作等工作。

【大数据应用开发（Python）】（中级）：根据业务需求和工作流程，主要使用 Python、MySQL、Hadoop、Linux 和 Pandas 等工具，完成大数据平台搭建、大数据采集、大数据处理与存储、大数据分析挖掘、大数据可视化等工作。

【大数据应用开发（Python）】（高级）：根据业务需求和工作流程，主要使用 Python、TensorFlow、Linux 和 Hadoop 等工具，完成大数据平台高级配置管理、大数据处理、大数据分析挖掘、模型优化与部署、文本挖掘、图像识别、语音识别等工作。

6.2 职业技能等级要求描述

表 1 大数据应用开发（Python）职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1. 数据管理	1.1 数据获取	1.1.1 能够根据业务需求基于 Excel 或 Power BI 实现数据录入与整理 1.1.2 能够根据业务需求基于 Excel 或 Power BI 实现 CSV 等文件的数据获取 1.1.3 能够根据业务需求基于 Excel 或 Power BI 实现网站数据获取 1.1.4 能够根据业务需求基于 Excel 或 Power BI 实现 MySQL 等关系型数据库的数据获取
	1.2 数据存储与查询	1.2.1 能够根据业务需求基于 MySQL 等关系型数据库实现 Excel、TXT 等文件存储与加载 1.2.2 能够根据业务需求基于 MySQL 等关系型数据库实现将数据存储为 Excel、TXT 等文件 1.2.3 能够根据业务需求基于 SQL 语句实现数据查询获取 1.2.4 能够根据业务需求使用 MySQL 进行数据库的同步、备份
	1.3 文件操作	1.3.1 掌握 Python 基本环境安装、Python 的基础语法、程序流程控制语句以及函数的使用 1.3.2 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的数据读取 1.3.3 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的数据写入 1.3.4 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的批量合并 1.3.5 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的批量拆分
2. 数据处理与分析	2.1 数据探索	2.1.1 掌握 Python 的列表、字符串、字典、元组、集合等数据结构的使用及常用操作 2.1.2 掌握 Python 面向对象编程 2.1.3 能够根据业务需求基于 Python 或 Excel 实现数据结构、数据分布的查看

		<p>2.1.4 能够根据业务需求基于 Excel 实现数据的频数分析</p> <p>2.1.5 能够根据业务需求基于 Excel 实现数据的集中趋势分析</p> <p>2.1.6 能够根据业务需求基于 Excel 实现数据的离散程度分析</p>
	2.2 数据处理	<p>2.2.1 能够根据业务需求基于 Excel 或 Power BI 实现数据缺失值、重复值、异常值的识别与处理</p> <p>2.2.2 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据类型转换</p> <p>2.2.3 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据合并</p> <p>2.2.4 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据提取、筛选等操作</p>
	2.3 数据分析	<p>2.3.1 能够根据业务需求基于 Excel 或 Power BI 实现同比、环比分析</p> <p>2.3.2 能够根据业务需求基于 Excel 或 Power BI 实现相关分析</p> <p>2.3.3 能够根据业务需求基于 Excel 或 Power BI 实现预测分析</p> <p>2.3.4 能够根据业务需求基于 Excel 或 Power BI 实现其他分析</p>
	2.4 数据可视化	<p>2.4.1 能够根据业务需求基于 Excel 或 Power BI 实现柱状图绘制</p> <p>2.4.2 能够根据业务需求基于 Excel 或 Power BI 实现散点图绘制</p> <p>2.4.3 能够根据业务需求基于 Excel 或 Power BI 实现饼图绘制</p> <p>2.4.4 能够根据业务需求基于 Excel 或 Power BI 实现其他个性化图表绘制</p>
3. 报表管理	3.1 报表制作	<p>3.1.1 能够根据业务需求基于 Excel 或 Power BI 实现数据透视表制作</p> <p>3.1.2 能够根据业务需求基于 Excel 或 Power BI 实现数据分析报表制作</p> <p>3.1.3 能够根据业务需求基于 Excel 或 Power BI 实现报表的图、表和文字的布局</p> <p>3.1.4 能够根据业务需求基于 Excel 或 Power BI 实现报表的美化</p>
	3.2 报表部署	<p>3.2.1 能够根据业务需求基于 Power BI 实现报表相关数据、文字描述和图表的发布</p> <p>3.2.2 能够根据业务需求基于 Power BI 实现仪表板的创建与设置</p> <p>3.2.3 能够根据业务需求基于 Power BI 实现</p>

		对发布后的数据共享 3.2.4 能够根据报表部署过程输出规范化的说明文档
	3.3 报表更新维护	3.3.1 能够根据业务需求定期更新报表 3.3.2 能够根据业务需求进行报表维护 3.3.3 能够根据业务变化持续迭代优化报表 3.3.4 能够及时发现和定位报表数据问题

表 2 大数据应用开发（Python）职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 平台管理	1.1 软件安装	1.1.1 能够根据业务需求实现 Linux 系统的安装与基础配置 1.1.2 能够根据业务需求实现 Python 集成开发环境的安装与基础配置 1.1.3 能够根据业务需求实现 MySQL 的安装与环境配置 1.1.4 能够根据业务需求实现 Linux 系统集群的搭建与基础配置 1.1.5 能够根据业务需求实现 Hadoop 相关组件或框架的安装与基础配置
	1.2 软件管理	1.2.1 能够根据业务需求实现 Linux 的状态监控与异常分析 1.2.2 能够根据业务需求实现 Python 集成开发环境的模块安装与加载 1.2.3 能够根据业务需求实现 PyCharm 的模块安装与加载 1.2.4 能够根据业务需求实现 MySQL 的状态监控与权限管理
	1.3 平台维护	1.3.1 能够根据业务需求定期监控各项运维数据 1.3.2 能够根据业务需求进行平台用户权限管理 1.3.3 能够根据业务需求在检测平台数据过程中通过数据分析发现业务问题并提出优化建议 1.3.4 能够根据业务需求保障平台的安全、稳定运行
2. 大数据采集与存储	2.1 大数据采集	2.1.1 能够根据业务需求实现完整的外部数据采集方案设计 2.1.2 能够根据业务需求基于 Request、lxml、re 等模块实现静态网页数据爬取 2.1.3 能够根据业务需求基于 Selenium 等框

		<p>架或逆向分析等相关技术实现动态网页数据爬取</p> <p>2.1.4 能够根据业务需求基于 Scrapy 等框架实现模块化的数据爬取</p>
	2.2 大数据清洗	<p>2.2.1 能够根据业务需求基于 Python 实现数据清洗和初步加工</p> <p>2.2.2 能够根据业务需求基于 Python 的列表、元组、数据框等数据结构实现数据结构化处理</p> <p>2.2.3 能够根据业务需求基于 Python 实现数据类型转换</p> <p>2.2.4 能够根据业务需要优化现有的数据清洗流程，制定数据清洗的流程，识别和规避数据清洗过程中的风险</p>
	2.3 大数据存储与管理	<p>2.3.1 能够根据业务需求实现将数据存储至 Excel、CSV、TXT 等文件</p> <p>2.3.2 能够根据业务需求实现将数据存储至 MySQL 数据库</p> <p>2.3.3 能够根据业务需求基于 Python 或 MySQL 实现数据关联、数据插入、数据修改等操作</p> <p>2.3.4 能够根据业务需求基于 Python 或 MySQL 等工具实现数据抽取</p>
3. 大数据分析 与挖掘	3.1 大数据处理	<p>3.1.1 能够根据业务需求基于 Python 实现数据异常值、缺失值和重复值的识别与处理</p> <p>3.1.2 能够根据业务需求基于 Python 实现数据变换、标准化、离散化等操作</p> <p>3.1.3 能够根据业务需求基于 Python 实现数据维规约、数值规约、特征构造等操作</p> <p>3.1.4 能够根据业务需求基于 Python 实现数据检索、数据排序等操作</p>
	3.2 大数据分析	<p>3.2.1 能够根据业务需求基于 Python 实现数据质量分析、关联分析、特征分析等操作</p> <p>3.2.2 能够根据业务需求基于 Python 实现线性回归、岭回归等回归预测</p> <p>3.2.3 能够根据业务需求基于 Python 实现决策树、逻辑回归、支持向量机、贝叶斯等分类预测</p> <p>3.2.4 能够根据业务需求基于 Python 实现 K 均值聚类、密度聚类、期望最大化聚类等聚类分析</p> <p>3.2.5 能够根据业务需求基于 Python 实现回归模型、分类模型、聚类模型的模型评估及参数调优</p>

	3.3 大数据可视化	<p>3.3.1 能够根据业务需求基于 Matplotlib、Seaborn、Pyecharts 等绘图模块或工具实现柱状图、散点图、饼图等图形绘制</p> <p>3.3.2 能够根据业务需求实现绘图参数的调整</p> <p>3.3.3 能够根据业务需求基于数据可视化结果总结出有效的分析结论，支撑市场决策</p> <p>3.3.4 能够根据业务需求基于数据可视化结果撰写相应的数据分析报告</p>
--	------------	---

表 3 大数据应用开发（Python）职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 平台管理	1.1 软件安装	<p>1.1.1 能够根据业务需求实现 Python 集成开发环境的高级配置</p> <p>1.1.2 能够根据业务需求实现 Hadoop 相关组件或框架的高级配置</p> <p>1.1.3 能够根据业务需求实现定制化环境搭建</p> <p>1.1.4 能够根据业务需求指导安装并制定平台使用规范与要求</p>
	1.2 架构管理	<p>1.2.1 能够根据业务需求实现架构选型</p> <p>1.2.2 能够根据业务需求实现架构设计与优化</p> <p>1.2.3 能够根据业务需求实现大数据平台到业务系统的端到端解决方案</p> <p>1.2.4 能够根据业务需求实现前后端开发团队的指导工作，促进项目落地</p>
	1.3 平台维护	<p>1.3.1 能够根据业务需求修改 Hadoop 相关组件或框架的高级配置</p> <p>1.3.2 能够根据业务需求定位平台报错位置</p> <p>1.3.3 能够根据业务需求对平台进行定期维护</p> <p>1.3.4 能够根据业务需求对平台数据进行冗余备份</p>
2. 大数据分析 与挖掘	2.1 大数据处理	<p>2.1.1 能够根据业务需求基于 Python 实现数据清洗、变换、合并、校验、特征分析等操作</p> <p>2.1.2 能够根据业务需求基于 Python 正则表达式实现文本处理</p> <p>2.1.3 能够根据业务需求基于 Python 实现文本分词、去停用词、词性标注与命名实体识别等操作</p>

		<p>2.1.4 能够根据业务需求基于 Python 实现文本向量化、文本特征计算、文本特征标准化等操作</p> <p>2.1.5 能够根据业务需求基于 Python 实现网络在线语料库获取等操作</p>
	2.2 模型构建	<p>2.2.1 熟悉线性模型、神经网络等分类与回归算法的原理并能够根据业务需求基于 Python 实现相应模型构建</p> <p>2.2.2 熟悉 K 均值聚类、密度聚类等聚类算法的原理并能够根据业务需求基于 Python 实现相应模型构建</p> <p>2.2.3 熟悉关联规则的算法原理并能够根据业务需求基于 Python 实现相应模型构建</p> <p>2.2.4 熟悉智能推荐、时序模式等算法原理并能够根据业务需求基于 Python 实现相应模型构建</p> <p>2.2.5 能够根据业务需求基于 Python 完成文本分类、聚类等文本挖掘任务</p> <p>2.2.6 能够根据业务需求基于 Python 的词典、主题模型等方法完成文本情感分析任务</p>
	2.3 模型优化与部署	<p>2.3.1 能够根据业务需求基于 Python 实现模型评价</p> <p>2.3.2 能够根据业务需求基于 Python 实现模型优化</p> <p>2.3.3 能够根据业务需求将模型部署至业务系统</p> <p>2.3.4 能够根据业务需求实现业务系统的模型维护</p>
3. 深度学习实战	3.1 大数据处理	<p>3.1.1 能够根据业务需求基于 Python 实现语音、图像、文本等数据的探索分析</p> <p>3.1.2 能够根据业务需求基于 Python 实现语音、图像、文本等数据的清洗</p> <p>3.1.3 能够根据业务需求基于 Python 实现语音、图像、文本等数据的标准化</p> <p>3.1.4 能够根据业务需求基于 Python 实现语音、图像、文本等数据的特征选择与构造</p>
	3.2 模型构建	<p>3.2.1 熟悉卷积神经网络、循环神经网络、长短时记忆网络等算法流程并能够根据业务需求基于 TensorFlow 实现相应模型构建</p> <p>3.2.2 能够根据业务需求基于 TensorFlow 完成图像处理、图像识别、语音识别、自然语言处理等任务</p> <p>3.2.3 能够根据业务需求基于 TensorFlow 搭建长短时记忆网络模型，完成文本分类、情</p>

		<p>感分析等任务</p> <p>3.2.4 能够根据业务需求基于 TensorFlow 搭建序列到序列模型完成语音翻译、中英文翻译等任务</p>
	<p>3.3 模型优化与部署</p>	<p>3.3.1 能够根据业务需求基于 TensorFlow 等深度学习框架实现模型搭建</p> <p>3.3.2 能够根据业务需求基于 TensorFlow 等深度学习框架实现模型评价与优化</p> <p>3.3.3 能够根据业务需求实现业务系统的模型部署</p> <p>3.3.4 能够根据业务需求实现业务系统的模型维护</p>

参考文献

- [1] GB/T 37696-2019 信息技术服务 从业人员能力评价要求
- [2] GB/T 36348-2018 信息技术 学习、教育和培训 虚拟实验 框架
- [3] GB/T 35295-2017 信息技术 大数据 术语
- [4] GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库
- [5] GB/T 5271.1-2000 信息技术 词汇 第1部分：基本术语
- [6] GB/T 5271.13-2008 信息技术 词汇 第13部分：计算机图形
- [7] GB/T 5271.31-2006 信息技术 词汇 第31部分：人工智能 机器学习
- [8] 教育部关于印发《职业教育专业目录（2021年）》的通知（教职成〔2021〕2号）
- [9] 《教育部关于公布2019年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2020〕2号）
- [10] 《教育部关于公布2020年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2021〕1号）