

金融大数据处理 职业技能等级标准

(2021年2.0版)

深圳希施玛数据科技有限公司 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）	3
6 职业技能要求	4
参考文献	10

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：深圳希施玛数据科技有限公司、深圳市金证科技股份有限公司、深圳前海信息技术有限公司、深圳市前海铂悦投资管理有限公司、上海瞰点科技有限责任公司、上海钒钛智能科技有限公司、深圳铸城科技有限公司、杭州轻影科技有限公司、安徽希施玛数据科技有限公司、深圳市丽海弘金科技有限公司、云天弈(北京)信息技术有限公司、深圳市链联科技有限公司、中国银宏有限公司、深圳市金多恩科技有限公司、Pure Era INC、安徽商贸职业技术学院、北京财贸职业学院、北京经济管理职业学院、北京培黎职业学院、北京青年政治学院现代管理学院、广西金融职业技术学院、河北金融学院、江苏财会职业学院、江苏经贸职业技术学院、南宁职业技术学院、山东女子学院、西安职业技术学院、邢台职业技术学院。

本标准主要起草人：沈立君、于久洪、胡君晖、胡增芳、白桦、张小春、韦智勇、林煜恩、李洁义、高健、吕林、潘进、许可、徐唐锋、俞伟、黄瀚铭、沈传河、田晓丽、苏跃辉、许庆建、张小裴、袁国强、董迪马、孙巍、朱洪海、周荣、陈工孟、陈星、江莲、李洵、何国杰、崔凯龙、刘威、汪健、穆旖旎、雷银彬、杨剑锋、郎朗。

声明：本标准的知识产权归属于深圳希施玛数据科技有限公司，未经深圳希施玛数据科技有限公司同意，不得印刷、销售。

1 范围

本标准规定了金融大数据处理职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于金融大数据处理职业技能培训、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 5271.1-2000 信息技术 词汇 第1部分：基本术语

GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库

GB/T 35295-2017 信息技术 大数据 术语

GB/T 35273-2020 信息安全技术 个人信息安全规范

3 术语和定义

GB/T 5271.1-2000、GB/T 35295-2017、GB/T 5271.17-2010/ GB/T 35273-2020界定的以及下列术语和定义适用于本标准。

3.1 大数据 big data

具有体量巨大、来源多样、生成极快、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]

3.2 关系数据库 relational database

数据按关系模型来组织的数据库。

[GB/T 5271.17-2010, 定义 3.2]

3.3 结构化数据 structured data

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

[GB/T 35295-2017, 定义 2.2.13]

3.4 非结构化数据 unstructured data

支持一个或多个应用领域，按概念结构组织的数据集合，其概念结构描述这些数据的特征及其对应实体间的联系。

[GB/T 35295-2017, 定义3.3]

3.5 元数据 metadata

关于数据或数据元素的数据(可能包括其数据描述)，以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

[GB/T 35295-2017, 定义 2.2.7]

3.6 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 35295-2017, 定义 2.2.9]

3.7 数据治理 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/T 5271.1-2000, 定义01.08.02]

3.8 量化投资 quantitative Investment

量化投资是指通过数量化方式及计算机程序化发出买卖指令，以获取稳定收益为目的的交易方式。

4 适用院校专业

4.1 参照原版专业目录

中等职业学校：会计、金融事务、统计事务、信托事务、电子商务、计算机应用、计算机网络技术、软件与信息服务等相关专业。

高等职业学校：金融管理、国际金融、证券与期货、互联网金融、投资与理财、信用管理、财务管理、信息统计与分析、计算机应用技术、大数据技术与应用等相关专业。

应用型本科学校：金融学、金融工程、投资学、经济学、金融数学、经济与金融、金融科技、计算金融、经济统计学、数据计算及应用、大数据管理与应用等相关专业。

4.2 参照新版教育专业目录

中等职业学校：会计事务、金融事务、统计事务、电子商务、大数据技术应用、软件与信息服务、计算机应用等相关专业。

高等职业学校：金融服务与管理、国际金融、证券实务、金融科技应用、财富管理、大数据与财务管理、统计与大数据分析、大数据技术、财税大数据应用、大数据与财务管理、大数据与会计、大数据与审计等相关专业。

应用型本科学校：金融学、金融工程、投资学、经济学、金融数学、经济与金融、金融科技、计算金融、经济统计学、数据计算及应用、大数据管理与应用等相关专业。

高等职业教育本科学校：金融管理、金融科技应用、保险、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、企业数字化管理、大数据工程技术、人工智能工程技术等相关专业。

5 面向职业岗位（群）

金融大数据处理职业技能等级证书，主要面向银行、保险、券商、信托等金融科技相关企业以及向互联网转型的政府、企事业单位数据采编、数据产品助理、数据分析师、数据产品经理量化策略研究员等岗位，主要完成数据统计汇总、数据采编、接口数据采集、数据清洗、数据结构化、数据存储、初级数据咨询服务、数据产品设计、轻量化的行业或企业微观数据分析、数据挖掘、数据可视化、量化投资应用等工作。

【金融大数据处理】（初级）：主要面向银行、保险、券商、信托等金融科技相关企业以及向互联网转型的政府、企事业单位数据采编、数据产品助理等岗位，主要完成数据统计汇总、数据采编、数据清洗、数据结构化、数据存储、初级数据咨询服务等工作。

【金融大数据处理】（中级）：主要面向银行、保险、券商、信托等金融科技相关企业以及向互联网转型的政府、企事业单位的数据采编、数据分析师、数据产品经理等岗位，主要完成接口数据采集、数据清洗、数据咨询服务、数据产品设计、轻量化的行业或企业微观数据分析等工作。

【金融大数据处理】（高级）：主要面向银行、保险、券商、信托等金融科技相关企业以及向互联网转型的政府、企事业单位的数据产品经理、数据分析师、量化策略研究员等岗位，主要完成数据标准化处理、数据存储、数据咨询服务、数据产品设计、数据挖掘、数据可视化、量化投资应用等工作。

6 职业技能要求

6.1 职业技能等级划分

金融大数据处理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。

【金融大数据处理】(初级): 根据业务需求, 在业务主管的指导下, 完成基本的宏观、行业、上市企业以及金融产品相关数据统计汇总、数据采编、数据清洗、数据结构化、数据存储、初级数据咨询服务等工作。

【金融大数据处理】(中级): 根据业务需求, 独立完成宏观、行业、上市企业以及金融产品相关接口数据采集、数据清洗、数据咨询服务、数据产品设计、轻量化的行业或企业微观数据分析等工作。

【金融大数据处理】(高级): 根据业务需求, 完成宏观、行业、上市企业以及金融产品相关数据标准化处理、数据存储、数据咨询服务、数据产品设计、数据挖掘、数据可视化、量化投资应用等工作。

6.2 职业技能等级要求描述

表 1 金融大数据处理职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1. 金融数据采集	1.1 数据合规认知	1.1.1 掌握宏观、行业、上市公司、金融产品常见数据指标, 并能说出其代表的含义; 1.1.2 熟悉并严格遵守数据采集相关法律法规; 1.1.3 能够有效性分析数据合规需求; 1.1.4 能够在合法的情况下进行数据库采集。
	1.2 数据源获取	1.2.1 能够找到宏观、行业、上市企业以及金融产品数据源; 1.2.2 熟悉各数据源的数据披露规律; 1.2.3 能够快速定位到宏观行业、金融产品以及上市企业数据出处及数据内容在公告中的具体位置; 1.2.4 能够对来自多个数据源的数据进行有效性分析并做出选择。
	1.3 金融数据手动采集	1.3.1 能够使用 PDF 编辑器、WPS 等工具将数据源获取到的资料转化为可编辑的数据; 1.3.2 通过质检工具对采集数据进行质检; 1.3.3 对质检结果发现的问题, 能够进行数据修复; 1.3.4 对所采集数据做简单的准确性比对检测。
	1.4 python 数据采集	1.4.1 熟悉互联网上常见数据终端种类; 1.4.2 掌握网站型互联网应用的常用开发语言、技术框架; 1.4.3 能够运用网页开发者工具来分析网页内容;

工作领域	工作任务	职业技能要求
		1.4.4 能够使用编译好的 python 采集模型，通过调参获取不同网站等互联应用的静态数据并进行抽取。
2. 金融数据清洗及存储	2.1 数据清洗	2.1.1 具备使用 Excel 或 Python 工具进行脏数据的初步清洗，处理常见的异常数据； 2.1.2 具备使用 Excel 或 Python 工具进行数据规范； 2.1.3 能够完成数据分解；能够对网站等互联网应用的静态数据进行准确拆解； 2.1.4 能够保证拆分后的数据属性都有明确的意义和目的。
	2.2 数据存储	2.2.1 能够根据互联网应用数据的特点合理选择适当的数据结构； 2.2.2 熟悉常见数据库的特点； 2.2.3 能将数据保存为不同类型文件； 2.2.4 具备将清理后的数据存储到目标库或数据表的能力。
	2.3 数据查询	2.3.1 能够正确安装 mysql 数据库； 2.3.2 掌握 SQL 基本数据查询命令的语法； 2.3.3 能够使用 SQL 数据库语句对所采集数据进行查询； 2.3.4 熟练使用 SQL 语句进行数据库错误排查。
3. 金融数据应用	3.1 数据咨询	3.1.1 能够根据咨询需求，规划要调研的数据指标，整理需求文档； 3.1.2 能说出金融数据的在不同业务场景中下的应用价值； 3.1.3 能够解答数据基础分析理论问题； 3.1.4 能够解答数据分析技术相关问题。
	3.2 数据可视化呈现	3.2.1 能够对所需呈现的数据，选择合适的图表模型进行呈现； 3.2.2 能够说出常见的几类图表的优缺点和使用限制； 3.2.3 能够使用常见的可视化工具，如 excel、Tableau、云图等快速绘图； 3.2.4 能够排版设计做出可视化图表报告。
	3.3 数据调研报告撰写	3.3.1 能够对调研数据进行简单业务分析和规律描述； 3.3.2 知道常见个股调研报告所需的数据指标，能够编写个股调研图文报告；

工作领域	工作任务	职业技能要求
		3.3.3 知道常见行业调研报告所需的数据指标，能够编写行业调研图文报告； 3.3.4 知道常见企业经营调研报告所需的数据指标，能够编写企业经营调研图文报告。

表 2 金融大数据处理职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. AI 自动化数据采集及清洗	1.1 数据规范	1.1.1 掌握数据源授权使用许可范围； 1.1.2 熟悉接口数据采集的技术标准和规范。
	1.2 接口数据采集	1.2.1 掌握接口的请求方式（get, post, 请求头, 请求体等）； 1.2.2 掌握基本数据结构（json）； 1.2.3 熟悉基本网络传输协议（http、https 等）； 1.2.4 了解基本的 Python 语句。
	1.3 文本数据处理	1.3.1 了解 OCR 识别技术的概念，NLP 自然语言处理技术的原理及使用场景； 1.3.2 能够使用文本数据采集工具对图片、pdf 文件等数据进行处理。
	1.4 数据清洗	1.4.1 掌握使用 python 语言进行数据去重的方法； 1.4.2 能够使用 python 语言的常见数据清洗手段对日期格式进行统一； 1.4.3 能够使用 python 语言对异常值及极端值进行处理； 1.4.4 掌握使用 python 语言对空值及空格进行处理的方法。
2. 数据库搭建及存储	2.1 数据库设计	2.1.1 能确定业务实施范围（业务用例）、用户行为（业务活动流程图）、输出 ER 图； 2.1.2 能够确保数据结构能够准确反映主题数据含义； 2.1.3 能够确保数据结构的完整性； 2.1.4 能够将 ER 模型转换为关系模型； 2.1.5 掌握根据关系模型确定数据库存储结构的方法。
	2.2 数据库搭建	2.2.1 掌握 mysql 搭建数据库的操作及方法； 2.2.2 能够基于方案设计，进行数据库搭建；

工作领域	工作任务	职业技能要求
		2.2.3 掌握 mysql 数据库的操作，具备数据库的管理能力； 2.2.4 掌握 mysql 数据库的操作，具备数据库的运维能力。
	2.3 接口数据入库存储	2.4.1 熟练掌握 sql 数据表的增、删、改、查操作； 2.4.2 能够使用 python 语言连接数据库； 2.4.3 能够使用 python 语言存储数据到数据库； 2.4.4 能够使用 python 语言在数据库中查找、提取及删除数据。
3. 金融数据分析建模	3.1 银行业数据分析	3.1.1 了解银行业通用信贷风控业务场景； 3.1.2 掌握特征选择的方法，并筛选出对违约状态影响最显著的指标； 3.1.3 能够根据案例数据，建立模型； 3.1.4 掌握使用机器学习预测及监控风险的方法。
	3.2 证券业数据应用	3.2.1 了解折线图、直方图、词云图等的使用场景； 3.2.2 掌握使用 python 语言根据市场情绪数据进行可视化绘图的方法； 3.2.3 能够根据可视化图表对市场情绪因子进行挖掘与分析； 3.2.4 能够根据证券市场微观结构，获取需要的因子； 3.2.5 根据选取因子，构建证券市场的分析模型； 3.2.6 能够根据所给案例数据，结合选取因子，对证券市场的情况进行预测。
	3.3 保险业数据分析	3.3.1 根据保险产品信息，明确营销问题，理解获取数据的业务逻辑； 3.3.2 掌握 python 中数据挖掘相关算法，如：关联规则计算及聚类算法分析等，对营销数据进行分析； 3.3.3 能够使用 python 中函数绘制出变量相关性热力图及用户画像图表。
	3.4 文本数据分析	3.4.1 掌握相关性分析的基本原理及方法； 3.4.2 能够使用 python 语言根据新闻词频数据及金融市场进行相关性分析。

表 3 金融大数据处理职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 数据治理	1.1 数据质量管理	1.1.1 能够根据需求构建数据评估体系，对数据进行规范性评估； 1.1.2 能够根据需求构建数据评估体系，对数据进行完整性评估； 1.1.3 能够根据需求构建数据评估体系，对数据进行及时性评估； 1.1.4 能够根据需求构建数据评估体系，对数据进行准确性评估； 1.1.5 能够确立数据质量改进目标、评估并制定改善计划，实施改进并评估改善效果。
	1.2. 数据安全 管理	1.2.1 掌握对隐私数据的加密脱敏、模糊化处理的方法，保障数据的安全运作； 1.2.2 能够提供数据库监控等多种数据安全 管理措施； 1.2.3 能够确保数据以精确和安全的方式得到备份在需要进行恢复时，能完整地装回系统中； 1.2.4 掌握使用特定的程序对服务器进行配置，为不同用户明确定义用户数据权限的方法。
	1.3 数据集成 管理	1.3.1 能够实现数据库与数据库之间的数据迁移； 1.3.2 掌握数据细化管理能力，创建有意义的数据细分； 1.3.3 能够创建灵活的数据标签； 1.3.4 能够将数据整合进行标准统一化的管理。
2. 金融数据分析	2.1 基于线性回归的收入预测	2.1.1 能够使用 Python 编写一元线性回归模型； 2.1.2 能够使用 Python 编写多元线性回归模型； 2.1.3 能够根据拟合度等指标评价模型训练及预测结果； 2.1.4 能够对模型预测结果进行可视化展示。
	2.2 基于逻辑回归的信用风险判定	2.2.1 能够用 Python 编写逻辑回归模型； 2.2.2 能够使用逻辑回归模型处理信用风险评估的实际问题； 2.2.3 能够根据准确率等指标评价模型训练及预测结果； 2.2.4 能够对模型预测结果进行可视化展示。

工作领域	工作任务	职业技能要求
	2.3 基于支持向量机的股票价格预测	2.3.1 能够用 Python 编写支持向量机模型； 2.3.2 能够使用支持向量机模型处理股票价格预测的实际问题； 2.3.3 能够根据准确率等指标评价模型训练及预测结果； 2.3.4 能够对模型预测结果进行可视化展示。
	2.4 基于 K 均值聚类模型的目标人群聚类分析	2.5.1 能够用 Python 编写 K 均值聚类模型； 2.5.2 能够使用 K 均值聚类模型处理人均收入分类的实际问题； 2.5.3 能够根据轮廓系数等指标评价模型训练及预测的结果； 2.5.4 能够对模型预测结果进行可视化展示。
3. 量化投资分析	3.1 时间序列分析	3.1.1 知道使用时间序列在金融业务上的常见几种实际应用场景； 3.1.2 会用 Python 的时间序列处理工具和数据算法； 3.1.3 能够进行常见时间序列的数组变换； 3.1.4 能够对时间序列分析结果进行绘图。
	3.2 量化策略搭建	3.2.1 掌握量化策略编写结构； 3.2.2 掌握量化策略数据 API 调用； 3.2.3 掌握并能开发经典量化策略，包括双均线策略、布林带策略及唐安奇通道策略等； 3.2.4 通过历史数据对量化策略进行回测，并能通过回测绩效指标评估量化策略。
	3.3 量化策略模拟交易	3.3.1 掌握创建模拟交易账户； 3.3.2 掌握配置模拟交易账户和量化策略； 3.3.3 掌握量化策略模拟交易的启停操作； 3.3.4 通过查看模拟交易数据和可视化界面，了解量化策略的实际运行情况。

参考文献

- [1] 中等职业学校专业教学标准（2017年版）
- [2] 高等职业学校专业教学标准（2019年版）
- [3] 《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）
- [4] 《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）
- [5] 《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）
- [6] 本科专业类教学质量国家标准（2018年版）
- [7] 《计算机科学技术名词》第三版（2018年版）
- [8] GB/T 5271.1-2000 信息技术 词汇 第1部分：基本术语
- [9] GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库
- [10] GB/T 35295-2017 信息技术 大数据 术语
- [11] GB/T 34960.5-2018 信息技术服务 治理 第5部分：数据治理规范
- [12] SJ/T 11615.1-2016 网络数据采集分析软件规范 第1部分：框架
- [13] GB/T 25058-2010 信息安全技术 信息系统安全等级保护实施指南
- [14] GB/T 35274-2017 信息安全技术 大数据服务安全能力要求
- [15] 《高频交易》，谈笑俊等译，机械工业出版社，2011年
- [16] 《量化投资与对冲基金》（2014年版）
- [17] 国家职业技能标准编制技术规程（2018年版）