



中华人民共和国国家标准

GB/T 41462—2022

基于文本数据的金融风险防控要求

Requirements for financial risks control based on text data

2022-04-15 发布

2022-11-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	4
5 整体框架	4
6 文本数据要求	4
7 预处理	5
8 信息抽取	5
8.1 概述	5
8.2 信息抽取整体框架	5
8.3 抽取内容及特征分析	6
8.4 抽取方法	7
9 数据表示	8
9.1 总体要求	8
9.2 数据表示评估	9
9.3 基于 RDFS 结构化表示	10
10 分析预警	13
10.1 数据清洗	13
10.2 建模方法	13
10.3 分析方法	14
11 用户交互	14
12 系统评估	15
12.1 原则	15
12.2 类别	15
12.3 评估方法	15
12.4 评估指标	16
13 安全防护	16
13.1 安全技术要求	16
13.2 安全管理要求	16
14 软硬件要求	16
14.1 硬件基本要求	16
14.2 软件基本要求	17

附录 A (规范性) 基于 AHP 的指标权重确定方法	18
附录 B (资料性) RDFS 结构化描述	19
附录 C (规范性) RDFS 表示具体技术流程	22
参考文献	23

前 言

本文件按照 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。

本文件由中国人民银行提出。

本文件由全国金融标准化技术委员会(SAC/TC 180)归口。

本文件起草单位：中国标准化研究院、中国银行业协会、北京理工大学、中国工商银行股份有限公司、中国农业银行股份有限公司、中国建设银行股份有限公司、中国工商银行股份有限公司深圳分行、华南师范大学、北京工业大学、中国科学院计算技术研究所、中国科学技术信息研究所、北京大学、上海对外经贸大学、中国金融电子化公司、中国人民银行太原中心支行、北京海致星图科技有限公司、聊城大学、江苏科技大学、广东外语外贸大学、北京师范大学、中版集团数字传媒有限公司、北京市科学技术情报研究所。

本文件主要起草人：曹馨宇、王海涛、刘涌、赵小林、郝天永、刘磊、王石、李宽、张漪漫、邢宸睿、刘耀、陈玉忠、曹存根、贾世军、穗志芳、刘亮亮、贾仰理、刘嘎琼、丁若尧、杨娟、聂大昕、严可、贺莉丽、李琪、薄舜添、邓琳莹、陈文俊、徐浩、陈全保、李辉、邬大港。

基于文本数据的金融风险防控要求

1 范围

本文件规定了基于文本数据金融风险防控的整体框架、文本数据要求、预处理、信息抽取、数据表示、分析预警、用户交互、系统评估、安全防护、软硬件要求。

本文件适用于金融相关的文本数据处理及金融信息的挖掘、抽取与分析。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 4754—2017 国民经济行业分类

GB/T 20269—2006 信息安全技术 信息系统安全管理要求

GB/T 20271—2006 信息安全技术 信息系统通用安全技术要求

GB/T 32319—2015 银行业产品说明书描述规范

ISO 21586:2020 金融服务的参考数据 银行产品服务说明描述规范 [Reference data for financial services—Specification for the description of banking products or services (BPoS)]

3 术语和定义

下列术语和定义适用于本文件。

3.1

学习 learning

一个生物学系统或自动系统获得知识或技能的过程，使它可用于改进其性能。

[来源：GB/T 5271.31—2006, 31.01.01]

3.2

概念 concept

为确定类别成员的抽象实体。

注：概念用于客体分类。

[来源：GB/T 5271.31—2006, 31.01.06]

3.3

语义 semantics

词或词组与它们的含义之间的关系。

[来源：GB/T 12200.1—1990, 4.1.2.12]

3.4

文本 text

文本数据 text data

以字符、符号、字、短语、段落、句子、表格或者其他字符排列形式出现的数据，旨在表达一个意义，其