ICS 35.240.40
CCS A 11

JR

中华人民共和国金融行业标准

JR/T 0263—2022

机器学习金融应用技术指南

Technical guidance on application of machine learning in financial services

2022－11－25 发布　　2022－11－25 实施

中国人民银行　　发布

目　　次

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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国人民银行科技司提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

机器学习金融应用技术指南

1 范围

本文件提供了金融业开展机器学习应用涉及的体系框架、计算资源、数据资源、机器学习引擎、机器学习服务、安全管理、内控管理等方面的建议。

本文件适用于开展机器学习金融应用的金融机构、技术服务商、第三方安全评估机构等。

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（包括所有修改单）适用于本文件。

GB/T 27910—2011 金融服务 信息安全指南

JR/T 0071.2—2020 金融行业网络安全等级保护实施指引 第2部分：基本要求

JR/T 0071.5—2020 金融行业网络安全等级保护实施指引 第5部分：审计要求

JR/T 0171—2020 个人金融信息保护技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

机器学习 machine learning

功能单元获取新知识或者技能，或通过已有的知识或技能改进其性能的过程。

[来源：GB/T 5271.31—2006，31.01.02]

3.2

监督学习 supervised learning

基于外部知识源的反馈以测试获得知识的正确性的学习策略。

[来源：GB/T 5271.31—2006，31.03.08，有修改]

3.3

无监督学习 unsupervised learning

无需基于外部知识源的反馈以测试获得知识的正确性的学习策略。

[来源：GB/T 5271.31—2006，31.03.09，有修改]

4 总体原则

机器学习在金融服务中应用的原则一般包括以下内容。

a) 普遍性。确认目标群中所有主体均能成功且一致地使用机器学习进行服务。

b) 不可否认性。涉及机器学习且已经发生的活动或事件不可被否认。

c) 可控性。主体对机器学习应用的使用范围、运行状态、访问权限等具备主动控制的能力。

5 体系框架

机器学习能为金融应用系统提供更丰富、更便利、更通用的智能化支撑服务，例如智能语音、自然语言处理、计算机视觉、生物特征识别、知识图谱等。基于机器学习的金融应用系统一般性体系框架见下图。

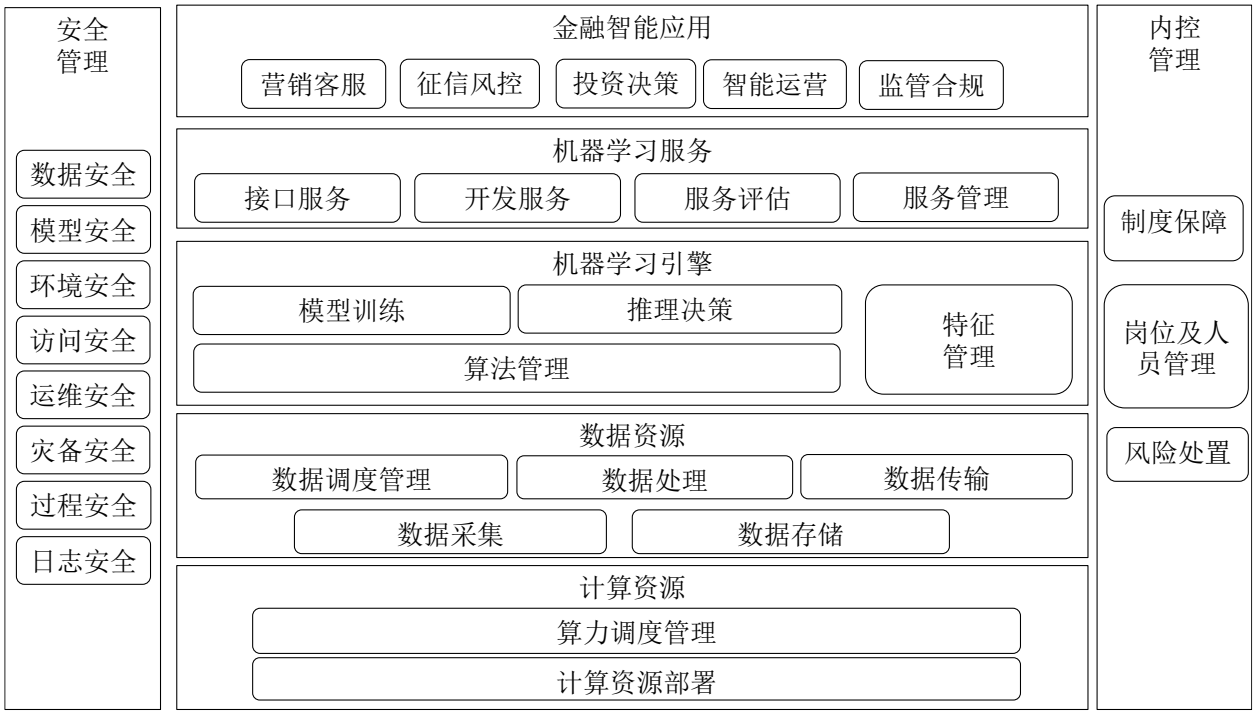

图　基于机器学习的金融应用系统一般性体系框架

基于机器学习的金融应用系统一般性体系框架包括计算资源、数据资源、机器学习引擎、机器学习服务、金融智能应用、安全管理与内控管理等。整个框架表明了机器学习金融应用可遵循的、适用于不同规模的一般性层次化抽象、结构及依赖关系。体系框架中包含如下部分。

a) 计算资源部分统筹管理多种可选用的系统底层计算硬件，例如中央处理器（CPU）、图像处理器（GPU）、张量处理器（TPU）、神经网络处理器（NPU）、可编程逻辑阵列（FPGA）和专用集成电路（ASIC）等，通过部署并调度管理各类计算资源，提供机器学习过程所需的算力支持。

b) 数据资源部分负责为机器学习的数据使用需求提供支撑，包含数据采集、数据存储、数据调度管理、数据处理、数据传输等过程。

c) 机器学习引擎部分为上层机器学习服务提供特征管理、算法管理、模型训练、推理决策等支撑。

d) 机器学习服务部分提供不同场景应用所需的基础通用服务框架，包含接口服务、开发服务、服务评估和服务管理等功能。

e) 金融智能应用部分提供面向最终用户、针对不同金融领域应用场景的综合服务，包含营销客服、征信风控、投资决策、智能运营、监管合规等，具体应用场景见附录。

f) 安全管理部分提供计算资源、数据资源、机器学习引擎、机器学习服务及金融智能应用整个体系的安全保障，包含数据安全、模型安全、环境安全、访问安全、运维安全、灾备安全、过程安全和日志安全等。

g) 内控管理部分为机器学习在金融场景的有效应用提供组织、方法和流程上的保障，包含制度保障、岗位及人员管理和风险处置等。

6 计算资源

6.1 计算资源部署

设计和部署底层计算资源时宜充分考虑其安全性、可靠性及可扩展性，需考虑的内容如下。

a) 宜整体考虑硬件、操作系统、上层计算服务组件的安全，确保底层算力的安全性。

b) 宜保证计算资源启动过程的安全性及上层应用的安全性，例如可通过逐级进行安全签名校验等方式确保固件的完整性和可信度。

c) 在实时性较强的机器学习金融应用系统中，宜保证计算资源的冗余、异构性和任务的可恢复性，从而保证业务连续性。

d) 宜保证架构的可扩展性，对计算资源进行虚拟化、池化管理，支持水平扩展，能提供按需（数据量、算法类型、实时性偏好等）分配计算资源的能力。

e) 宜定期开展底层计算资源安全性的审核工作，监控设备运行状态、资源使用情况，发生异常情况及时告警。

f) 宜根据数据密集程度、格式、数据量方面的差异合理分配计算资源，以达到理想的实时处理效果。

g) 宜具备对不同硬件资源的兼容能力，例如服务器、一体机、边缘计算节点、计算集群和云基础设施等。

h) 宜提供异构资源管理调度功能，屏蔽下层硬件的异构复杂性，提供统一的计算资源服务接口。

6.2 算力调度管理

算力调度管理过程涉及多模块间的协同工作，宜重点关注过程中的数据隐私、数据安全等问题，需考虑的内容如下。

a) 在不同任务或计算节点间切换时保证数据处理过程的保密性，避免因调度过程导致数据泄露等问题。

b) 在人机协同模式中，宜保证主机安全并使用完整的加固流程，例如严格裁剪服务、进行网络端口扫描、开启主机入侵检测和漏洞管理服务。

c) 在云化协同模式中，宜开展周期性安全指数评估工作，部署信息防泄露监控工具，保证金融行业数据中心运行环境位于高安全区域。

d) 宜定期检查或评估数据连接通道的安全性和可靠性。

e) 宜根据场景和运算需求分解应用负载，并分配任务执行的优先级。

f) 宜按需使用多处理单元，实现多芯片、多核、多机、多上下文并行。

7 数据资源

7.1 数据采集

采用机器学习的金融应用系统的数据采集过程需考虑的内容如下。

a) 对数据采集的来源和方式、数据范围、数据形式、数据状态等方面进行说明。

b) 对数据采集的环境、工具、技术以及校验方法等采取必要的管控措施，保证所采集数据及其标记的完整性、真实性和一致性。
c) 数据采集后，宜对采集数据进行初步清洗，包括去除重复及错误数据、补充残缺数据等。
d) 宜记录数据采集过程中的活动及其责任人，保证采集活动的可追溯性和可审计性。
e) 个人金融信息数据和金融业务敏感数据的采集过程宜符合 JR/T 0171—2020 相关要求。

7.2 数据存储

采用机器学习的金融应用系统的数据存储过程需考虑的内容如下。
a) 在必要时，宜对采集或应用生成的数据进行持久化存储，以便后续用于模型的训练及校验。
b) 数据的存储宜设置存储时限，并满足时间最小化原则，即为满足机器学习目的所必需的最短时间。
c) 在超出数据存储时限后，宜立即对所存储的数据进行处理（例如删除、销毁、脱敏等）。
d) 宜支持结构化存储方式（例如关系型数据库）和非结构化存储方式（例如键值数据库、图数据库等）。
e) 个人金融信息数据和金融业务敏感数据的数据存储过程宜符合 JR/T 0171—2020 相关要求。

7.3 数据处理

采用机器学习的金融应用系统的数据处理过程需考虑的内容如下。
a) 宜采用规范化的数据处理过程和数据校验机制，保证数据处理结果的一致性和准确性。
b) 宜支持数据聚合功能，即对分散的数据进行融合或拼接。
c) 宜对数据处理过程进行管理和记录，保证数据处理活动的可追溯性和可审计性。
d) 个人金融信息数据和金融业务敏感数据的委托处理和加工处理过程宜符合 JR/T 0171—2020 相关要求。

7.4 数据传输

采用机器学习的金融应用系统的数据传输过程中，对适用于 JR/T 0171—2020 相关要求的个人隐私及业务敏感数据，宜建立相应的数据传输安全策略和规程并采用有效的技术手段和控制措施，确保数据在传输过程中的保密性、完整性和可用性。

7.5 数据调度管理

采用机器学习的金融应用系统的数据调度管理过程需考虑的内容如下。
a) 宜支持数据宿主选择，即选择数据服务过程中承接数据或驱动调度的节点。
b) 对于智能金融应用场景中高频次的实时推理决策需求（例如事中不当操作检查），宜使用适当拥塞控制策略以确保整个系统处理链条的健康状态。
c) 宜提供弹性扩展功能以增强系统整体吞吐量并分摊数据服务压力。
d) 宜支持执行算子抽象技术，将数据流控制算法、数据获取算法等功能抽象为统一数据服务接口，并遵循标准通信协议，避免跨机构金融协作活动场景中出现数据格式、协议不一致等情况。
e) 宜支持数据虚拟化和基于规则的报警机制（例如日志审计）等，确保数据服务的整体健壮性。

8 机器学习引擎

8.1 特征管理

金融应用系统中机器学习引擎的特征管理需考虑的内容如下。

a) 宜保证特征的保密性，使特征在产生、传输、处理和存储的各个环节中不被泄露给未授权的个人和实体。
b) 宜保证特征的完整性，即特征在传输过程中不被篡改、破坏和插入，不发生延迟、乱序和丢失的情况，特别是采用分布式存储方式时，确保内容的一致性。
c) 宜保证特征的可用性，使特征可被已授权的其他机器学习子系统或实体使用。
d) 根据算法的需要，宜支持自动及手动选择与构建特征。
e) 宜支持离散特征、连续特征、时序特征、组合特征的抽取。
f) 宜支持显式特征构造方法，以增强特征的可解释性。
g) 宜保证离线特征与在线特征的一致性。

8.2 算法管理

在基于机器学习的金融系统核心业务中，例如核心交易系统、清结算系统、量化交易系统等，算法设计者宜提高算法的鲁棒性，增强安全性，需考虑的内容如下。

a) 当训练数据中有恶意数据破坏原有训练数据分布时，宜能区分和识别恶意数据，防止模型精度降低。具体对策可包括增加算法参数、丰富特征库、优化权重比例等。
b) 在算法设计过程中，宜综合考虑安全性、泛化性能和算法开销，合理权衡算法的安全性、实用性和性能，以更好地满足实际应用需求。
c) 宜采用集成学习、模型融合等提升手段集成不同的决策方法，以增强模型泛化能力，更好地适应未知数据分布情况。

8.3 模型训练

8.3.1 训练准备

金融应用系统中机器学习引擎的训练准备需考虑的内容如下。

a) 宜根据金融应用场景需求建立模型选择策略，策略选择依据包括但不限于模型复杂度和可解释性、原始数据及中间数据规模和维度、存储与计算资源成本、金融业务诉求精度和准确率、模型结果的时效性以及潜在的外围干扰项。
b) 宜根据金融应用场景需求定义约束条件和评价指标，并将其转换为对应监督学习或无监督学习的性能指标和评价函数，具体内容如下。
   ——监督学习评价指标，例如混淆矩阵、查准率、查全率、准确率、置信度、误差平方和、决定系数、对数似然损失函数、受试者工作特征曲线（ROC）的曲线下面积（AUC）值等。
   ——无监督学习评价指标，例如方差、还原误差、置信度、困惑度、类间距离、类内距离等。
c) 当数据量相对较少时，宜采用交叉验证，例如简单交叉验证、留一交叉验证等。

8.3.2 训练过程

金融应用系统中机器学习引擎的模型训练过程宜根据金融应用场景需求选择合适的训练方法，需考虑的内容如下。

a) 宜使用特征工程方法提升模型效能，具体内容如下。
   ——引入行业专家筛选优选特征，借助特征可视化提升场景描述效率。
   ——强化原始特征数学统计描述并提升特征表达能力，并通过特征组合或拆解提取特征的相关性，调整数据特征以增强特征可分析性。
   ——构造适用于训练学习过程和结构的特征以便机器学习算法使用。
   ——通过坏样本发现易被忽视的特征。

——采用上下采样或数据权重调整等方法进行数据均衡化处理。

——挖掘数据异常点和孤立点并采用数据清洗和数据变换等方法降低数据噪声。

b) 宜根据场景需要提升训练模型的稳定性，具体内容如下。

——可通过数据增强技术和数据扩展技术，提高模型泛化、抗噪能力。

——对于模型训练过程本身，宜使用剪枝、部分激活、批正则化或权重衰减等方式降低算法对训练数据的过拟合风险。

——在模型参数设置阶段宜根据数据分布特点进行合理初始化，特别是对于无监督学习方法，以提升模型稳定性、加快收敛速度。

——根据应用场景特点，对模型进行定期评估更新以适应用场景的变化，并对模型更新过程作出规范，使用模型再训练、在线学习、迁移学习等方式保证模型持续有效。

c) 宜根据应用场景数据规模优化训练效果，具体内容如下。

——对指标不敏感的场景，可通过调整学习率以加快模型训练收敛。

——对于较复杂或规模较大的业务网络，可使用残差网络等方式解决梯度衰减问题。

——针对小批量样本，可扩充数据集，保证训练集的数量级不小于模型的复杂度。

——通过调整模型结构和模型正则化，降低模型复杂度，保证模型的复杂度不大于训练集的数量级。

——在监督学习训练数据集中引入对抗样本，通过对抗攻击构建高容忍模型，提高鲁棒性。

8.4 推理决策

推理决策宜满足金融应用场景需求，需考虑的内容如下。

a) 宜保证模型的可用性，对样本数据有较好的容忍度，在极端情况下，保证模型可以正常返回结果，供系统进行决策处理。

b) 宜保证输入样本数据及输出返回结果的保密性，确保不被未授权用户非法获取。

c) 宜保证输入样本数据及输出返回结果的完整性，确保不被非法篡改。

d) 宜保证关键性场景中模型的可解释性，从业务建模、参数设置和样本选择等多方面提升模型可解释性。

e) 宜使用模型压缩或者剪裁提升推理速度，并对使用的压缩、蒸馏、裁剪方法记录备案。

f) 从多方面提升推理决策服务的可用性和效率，主要内容如下。

——具备任务按需调度能力，根据模型、业务特点确定计算节点和存储节点。

——统一管理中心集群与边缘节点，使边缘业务就近调度到边缘设备执行。

——具备任务跨集群调度与多级资源拉通共享能力，可将本地任务调度到另一个集群中计算。

9 机器学习服务

9.1 接口服务

金融应用系统中机器学习服务的接口服务需考虑的内容如下。

a) 宜制定机器学习服务接口调用的技术规范，并确保外界无法通过接口获取相关隐私或者敏感数据。

b) 对于机器学习服务接口的定义，宜尽量缩小输入输出的数据范围，保证输入参数和输出数据没有冗余，防止额外的数据通过接口泄露。

c) 机器学习服务接口调用宜做好数据参数检查，防范数据库注入攻击等风险。

d) 算法接口宜支持对算法的能力信息进行展示，包含算法名称、算法运行环境信息、算法运行能力。

e) 宜支持对算法资源进行管理，包含算法资源增加、算法资源修改、算法资源删除、算法资源查询、算法资源下载等接口。
f) 宜对算法接口处理结果返回状态进行明确定义，例如处理成功、处理失败、未匹配等。
g) 宜对接口返回内容进行明确定义，例如分类标签范围、概率值区段等。
h) 算法接口宜支持多种市场主流工具和模型文件格式。

9.2 开发服务

金融应用系统中机器学习服务的开发服务需考虑的内容如下。
a) 宜提供数据集管理、半自动化标注、模型超参数自动调优、自动机器学习、可视化建模与评估等功能，减少外接组件数目和配置工作量，降低兼容性风险。
b) 宜提供图形化工作流编辑界面，快速构建应用处理逻辑，减少金融系统复杂场景应用的开发维护工作量。
c) 宜在训练过程中，通过适当的自动化模型质量评估手段（含可视化），来反馈模型的质量，提升模型选择及构建的效率。
d) 宜兼容不同架构的机器学习框架及不同数据源。
e) 宜在模型开发、训练、构建、数据处理等过程中，为开发人员或生产运营人员提供机器学习服务、各类资源的使用情况和可用性的监控手段。
f) 宜支持学习训练、数据处理等过程中各类计算资源的扩容，以适应不同数据规模和模型复杂度的计算需求。
g) 宜支持模型训练和模型预测服务中所需要的各种周边辅助算子，包括数据分析、隐私集合求交等多种数据预处理算子。

9.3 服务评估

金融应用系统中机器学习服务的服务评估需考虑的内容如下。
a) 宜具备对机器学习的任务进行状态跟踪与记录的功能。
b) 宜构建完整的机器学习金融应用模型上线与评估机制，包含效果评估、性能评估、安全评估、对比测试、上线流量控制等机制。
c) 机器学习金融应用新模型上线后，宜定期观测新模型在线上的效果、检测模型在测试与线上环境下的效果和性能差距并进行迭代训练。
d) 机器学习金融应用模型上线宜采用对比测试和灰度发布机制，使得上线过程能够平滑过渡,且具备版本回溯功能。

9.4 服务管理

金融应用系统中机器学习服务的服务管理需考虑的内容如下。
a) 机器学习金融应用系统开发宜做好模块分离，包括但不限于降低应用层模块、算法层模块之间的耦合。
b) 宜对机器学习服务进行封装，降低使用方的接入成本。
c) 宜支持数据源水平切分及垂直切分、模型训练和模型预测等服务。
d) 宜支持对机器学习分布式任务的分解与调度。

10 安全管理

10.1 数据安全

金融机构从数据的采集、传输、存储、使用、删除、销毁等方面建立全生命周期防护措施，需考虑的内容如下。

a) 宜符合 JR/T 0171—2020 和 GB/T 27910—2011 的相关要求，确保机器学习应用过程中的个人隐私数据安全和业务敏感数据安全。
b) 宜制定机器学习数据采集过程中所涉及的硬件投入计划、源数据量、源数据格式、数据来源的采集结果评定要求。
c) 宜对机器学习金融应用模型的测试数据进行脱敏、权限控制等安全保护处理，测试数据要隔离存储，并基于线上数据定期更新、扩充。

10.2 模型安全

金融应用系统中的机器学习服务宜确保模型在建模、部署、应用、撤销等环节的安全，制定相应的策略来应对模型探测与窃取，需考虑的内容如下。

a) 在模型构建阶段，宜采取加密等措施保障模型建模过程安全，防止模型被窃取或恶意篡改，同时加强训练数据安全管理，在维持模型规模和准确率的同时，选择并生成表面更平滑的、对扰动不敏感的模型，防止异常数据、对抗攻击导致的模型错误。
b) 在模型部署阶段，宜确保模型到业务系统的传输安全，并具备自动检测机制，确保由专人部署到目标系统。
c) 在模型应用阶段，宜采用访问控制、探测频率限制等措施防范恶意样本攻击、异常高频探测、访问权限越界等安全风险，确保攻击者在探测模型时获取尽可能少的反馈。
d) 在模型撤销阶段，宜注销模型中与业务相关的所有敏感信息，记录撤销模型编号，并将其加密存储管理。
e) 宜对机器学习金融应用的模型开发过程或模型优化过程制定流程管理规范与安全评估方案。
f) 机器学习金融应用模型的训练、测试、预测环境宜使用统一的模型规范，后续变更时以训练环境所依赖的框架和软件为准。

10.3 环境安全

对金融应用系统中的机器学习服务运行环境进行安全管理，需考虑的内容如下。

a) 宜保障机器学习金融应用全流程设备运行环境安全，包括但不限于数据采集、模型训练、模型部署、模型应用时设备所在环境。
b) 机器学习金融应用全流程运行环境的物理安全、主机安全、网络安全宜符合 JR/T 0071.2—2020 的相关要求。
c) 宜定期检查机器学习金融应用全流程运行环境的开源框架、第三方依赖库，保障机器学习所依赖框架、库函数的安全。
d) 宜将机器学习模型训练环境与模型应用环境隔离，训练脚本、部署脚本单独加密存放。
e) 宜保证运行环境相关配置信息的数据安全，防止泄露或恶意修改。

10.4 访问安全

对金融应用系统中的机器学习服务进行访问管理，需考虑的内容如下。

a) 机器学习金融应用系统的网络、主机、应用访问控制管理，宜符合 JR/T 0071.2—2020 的相关要求。
b) 宜结合业务需求定义机器学习金融应用系统的用户访问控制策略，主要包括以下内容。
   ——宜对数据和计算资源进行定义和分类，并根据用户权限控制数据资源访问过程和计算资源分配过程。

——宜根据用户角色赋予相应的管理权限，包括但不限于数据工程师、算法工程师、业务分析师等，并建立责任追踪机制。
——宜建立用户身份鉴别机制，有效辨识用户角色，防止恶意用户窃取或篡改数据和模型。
——宜定义重要事件并记录，例如模型参数变动、推理参数变动、异常样本等，并根据用户的唯一标识对其访问、写入数据和模型等行为进行记录。
——宜提供对机器学习服务接口调用者的权限管理，通过访问控制、行为监控、日志记录等安全措施，防止未经授权的访问。
——宜具备重要页面水印功能，在展示重要数据相关页面时可设置页面水印标识，防止截图、拍照等手段窃取信息。

10.5 运维安全

对金融应用系统中的机器学习服务进行安全管理，需考虑的内容如下。
a) 宜建立机器学习金融应用的模型管理、回退和迭代的安全管理机制。
b) 宜建立机器学习金融应用的模型反馈和数据收集机制，以便后续对模型进行优化更新，提升模型性能和安全水平。
c) 若机器学习金融应用需提供在线服务，宜建立相应机制，持续对服务进行安全监控，防止模型发生异常。
d) 根据业务特征，宜建立机器学习金融应用的模型评估机制，定期或在运行外部环境发生变化时对模型安全进行评估。
e) 宜保持机器学习金融应用的模型评估、优化和重新部署上线等方面的开发过程与控制过程一致。
f) 宜保证运维工具与模型运行系统相互独立。
g) 机器学习金融应用宜记录运维管理过程中的重要活动及其责任人。
h) 宜对机器学习金融应用开发过程中的计划、交付、关键路径、风险及其干系人进行记录。

10.6 灾备安全

金融应用系统中的机器学习服务宜按照 JR/T 0071.2—2020 的相关要求进行灾备管理，提供合理的保障措施，确保机器学习金融应用系统发生灾难或者系统错误时业务的连续性。

10.7 过程安全

金融应用系统中的机器学习服务在技术实施过程中宜满足 JR/T 0071.5—2020 的相关要求，实施内部审计，保留运行依据并开展定期审查。

10.8 日志安全

对金融应用系统中的机器学习服务进行日志管理，提供审计依据，需考虑的内容如下。
a) 机器学习金融应用系统宜记录系统运行日志，主要包括以下内容。
——机器学习金融应用系统信息，包含使用者所针对的数据类型及来源、所选取的算法及相关配置、所调用和存储的相关模型参数、所操作的软硬件环境信息、使用者的名字等。
——记录机器学习金融应用系统相关设备生命周期管理事件，包含设备收据、存储设备的增加或移除、设备用途、设备卸载、有关设备使用和维修的指派、设备停用等。
——记录安全敏感事件，包含安全敏感事件的读取、写入、删除、变更和对机器学习金融应用系统或者任何组件的访问。
——记录事件日志，包含日期和时间、序列号、类型、源（终端、端口、地点、客户等）、制作日志的实体标识、日志等级等信息。

b) 机器学习金融应用系统的事件日志宜定时进行归档，采用签名加密等方式进行存储，周期性审查日志的完整性，并对异常、未授权或者可疑活动进行识别和跟踪。

11 内控管理

11.1 制度保障

开展机器学习金融应用的金融机构宜制定完善的管理制度，需考虑的内容如下。

a) 宜建立机器学习开发、应用各流程的操作合规性审计制度，建立违反相关管理制度的惩罚和豁免制度。

b) 宜制定面向机器学习应用的培训教育管理制度、档案制度，面向干系岗位制定培训教育计划，内容包括但不限于专业技术技能、应用安全制度、风险及处置程序等，并定期检查考核干系岗位员工掌握情况。

c) 宜建立应急保障制度、第三方库管理制度、外包服务引入制度、第三方数据或产品引入制度。

d) 宜建立第三方软件使用审查控制机制，主要包括以下内容。

——设置专人负责第三方软件的遴选，并设置专人对遴选结果和过程进行审核。

——对第三方软件的获取、保存、安装、使用、升级、回退、变更、卸载进行记录，详细程度宜符合金融机构内容审计要求，并实现记录分离保管和限制以上操作实施者对记录的访问。

——设置专人分析第三方软件的使用许可证等法务事项，避免不当获取及使用造成的侵权。

——对第三方软件的保存配备安全措施，防止第三方软件被非授权修改和配置。

——设置专人负责第三方软件的升级或回退，宜在实验环境中进行功能、性能、安全指标测试，符合金融业务要求后再对线上系统进行变更。

——宜在实验环境预先对第三方软件进行卸载测试，分析其去除的依赖组件及对系统的影响，再在线上系统实施，降低第三方软件卸载对线上业务的影响。

——检测第三方软件供给单位披露的关于该第三方软件的漏洞和相应补丁，在发现漏洞时，确定处置办法并经技术评议后，由专人实施。

e) 宜建立外包管理制度，主要包括以下内容。

——建立总体规划，明确计划周期内外包事项的目的、方式、时间控制、责任人等。

——建立外包管理组织架构，明确机器学习应用外包业务归口管理部门，制定相关管理政策，保证外包服务有协议、服务合同和监督机制约束。

——完善外包服务商准入和审查、评估及风险管理制度，对外包服务商的资质、人员管理、信息安全管控、知识产权、质量保证进行审查，对重要的外包服务商进行尽职调查。

——搭建外包风险防控体系，对外包风险进行分析评估，针对外包问题引发的信息系统中断、训练样本数据等敏感信息泄露、信息安全等问题，采取风险缓释、转移、规避等措施。

——建立全生命周期管理的外包机制及外包安全管理效果衡量机制。

——对外包事项，做到事前评估、事中监控、事后评价。

11.2 岗位及人员管理

开展机器学习金融应用的金融机构宜做好机器学习开发和应用中的人员管理工作，需考虑的内容如下。

a) 宜明确管理第一责任人，对机器学习应用的合规、稳定、性能等方面全面负责，同时可指定专人负责日常管理工作。

b) 宜成立领导小组或者管理办公室，负责具体管理实施工作，包括政策、法律和法规的贯彻落实，机器学习金融应用管理总体策略、管理规范和技术规范等的制定和推广。

c) 对各核心信息安全岗位配备管理员，具体岗位宜考虑以下内容。

——设立系统运维岗，保障机器学习应用的运行稳定性、服务能力等级。

——设立模型评审岗，对机器学习模型的可用性、可解释性、精确性、算法执行效率等指标进行评估和改进。

——设立合规管理岗，对机器学习系统开发和应用过程中的合法、合规性进行审核。

——设立相关培训教育责任归口管理岗。

d) 宜制定每个岗位的安全培训要求、培训程序、再培训周期和再培训程序，并组织相关人员按照岗位要求进行安全培训。

e) 宜与相关人员签署保密协议，主要包括以下内容。

——训练及测试数据。

——用户的隐私信息。

——算法模型的设计、部署、运行等信息。

f) 宜制定纪律处理程序，对违反安全要求的人员采用此程序进行处理，并采取有效措施保证相关人员合同终止时安全性不受影响。

11.3 风险处置

开展机器学习金融应用的金融机构宜明确风险管理的目标、范围、人员、评估方式、评估结果形式等内容，对其面临的机器学习应用风险进行管理，需考虑的内容如下。

a) 宜制定风险识别方案，形成覆盖风险来源、影响区域、风险类型、风险致因等识别项的规范。

b) 宜制定风险分析方案，依据风险识别结果形成风险影响后果评估准则，包括正面后果、负面后果及其发生的概率。

c) 宜制定风险评价方案，依据风险分析结果形成待处理风险的优先级，并提供参考处理规程。

d) 宜制定风险处理方案（预案），依据风险评价结果确定风险处理方案，并对方案进行实施和结果反馈。

e) 宜建立机器学习金融应用的技术、管理风险的应对措施，主要包括以下内容。

——如已明确机器学习金融业务组件存在模型、数据非授权篡改等情况，需立即报告，分析潜在危害范围及程度，形成风险评价结论，并申报审批。按照预案，对风险源头组件，进行替换或监控。

——如已明确机器学习金融业务组件存在决策失当等情况，需立即报告，分析潜在危害范围及程度，委托专人检测、引导舆情，同时实施技术分析，实施纠正、干预预案，宜包含再训练，组件代替等。

——如已明确机器学习金融业务组件存在模型、数据可能被窃取等情况，需立即报告，分析安全漏洞，按照预案实施安全加固措施。在完成前，对核心金融业务决策组件实施附加的访问控制策略。

——如已明确机器学习金融业务组件遭到外部攻击，需立即采取访问控制策略，阻隔攻击源对业务组件的访问，同时对攻击手法展开分析、研究，实施技术应对方案。在完成之前，宜暂停该业务组件在金融业务逻辑中继续服务或实施附加的访问控制策略。

f) 宜制定安全风险全生命周期管理方案，明确风险识别、风险分析和评估的频率，制定应急风险评估的触发条件，明确风险评估结果的验证方法和规范。

g) 宜建立健全相关风险补偿机制，在客户因相关机器学习金融应用服务受损及投诉时，明确管理及技术责任和相应补偿方法，切实保障客户合法权益。

附　录
（资料性）
金融领域机器学习应用场景

1　征信与风控

1.1　信用评估应用

主流的信用评估方法主要包括专家评分、评分卡评分等，受主观因素影响较大，金融机构主要使用个人征信报告中的分数、工资收入等少量传统指标。在大数据时代，通过使用机器学习，金融机构可以利用海量多维素数据建模，综合评定信用评分。基于机器学习的信用评估信息，是传统金融企业信用评估手段的有效补充。

1.2　授信融资应用

金融机构在实践普惠金融的过程中，部分贷款产品基于抵押数据、历史交易还款数据等信息进行授信，存在只能服务于存量客户的局限性；另一部分贷款产品虽然覆盖到了长尾人群，但存在服务价格较高的问题。

利用机器学习和海量第三方数据，可建立风险评估和授信模型，实现为更多客户授信，让更多人建立起个人信用体系的效果。同时，使用机器学习，可大幅提高金融服务效率，实现授信结果快速、贷款系统自动审批、运营成本降低、服务价格降低、普惠金融全面实施的显著成效。

1.3　反欺诈应用

反欺诈场景主要存在于银行、证券和保险等企业中。传统的交易欺诈侦测一般采用事后监控与专家分析相结合的形式，通过总结形成相应的专家规则，从而对可疑交易进行警告。虽然传统方法易于理解，但在准确性和时效性上仍有较大的提升空间。随着时间推移，技术发展日新月异，欺诈分子的手段和技术也不断迭代更新，使用传统方法难以满足新形势下反欺诈的需求。通过使用机器学习与多层神经网络技术，结合客户多维度数据，构建基础特征、时序特征、趋势特征、统计特征进行建模，发现隐藏在历史数据中的规律，通过模型实现欺诈行为精准识别、欺诈行为快速响应、复杂欺诈场景深入研判以及新型欺诈形式识别。

2　营销与客服

2.1　智能问答应用

在客户服务场景中，目前大多通过人工服务以在线或者电话沟通的方式解决客户的问题，存在客户服务运营成本高和客户体验差的问题。利用自然语言处理技术，结合智能语音技术和知识图谱技术实现智能服务机器人技术，以降低服务成本，提升客户体验。

2.2　精准营销应用

应用机器学习建立的企业级和产品级的营销推荐引擎，在得到授权后使用客户属性信息、客户资产信息、客户交易信息、客户消费行为信息、客户存贷款信息等数据，结合产品购买数据，并借助机器学习建立精准营销模型，形成千人千面的营销服务体系，洞察客户心理、了解客户需求。

2.3 智能投顾应用

智能投顾结合投资者提供的风险偏好、投资收益要求、投资风格信息、投资组合产品信息、产品收益信息，使用机器学习进行建模，为用户提供投资决策信息参考，并随着金融市场动态变化对资产组合及配置提供改进的建议。

通过对投资者进行简单的风险问卷调查，智能投顾应用得到投资者风险偏好、预期收益、投资风格、投资金额等基本信息，利用机器学习技术智能匹配资产配置模型，为客户定制化产生匹配的投资组合。

2.4 智能保顾应用

智能保顾场景中，保险公司通过知识图谱技术从保险产品信息中自动批量抽取相关数据，经过数据清洗和计算后，构建保险知识库，为客户及保险人员提供核保、投保等保险顾问服务，并可根据客户需求提供保险购买建议，提高保险投保阶段及运营过程的效率。

3 投资与决策

3.1 舆情分析应用

金融市场产品价格及成交行情通常受到多方面市场因素及突发事件的影响，为了提升应用模型对于市场整体形势的理解能力，可利用自然语言处理技术，深度挖掘从不同市场媒介及社交网络获取的金融舆情信息，掌握市场中不同因素及事件的发展态势，提升对金融市场未来趋势预测的准确率。

3.2 智能投研应用

使用机器学习对数据进行获取、挖掘、分析，可实现标的搜索、自动生成研究报告等功能。金融产品中存在大量的非结构化文本及表格数据，例如财报、研报、公告、合规报告等，其中蕴含丰富的业务价值，传统上往往需要大量人力劳动进行解读分析，通过自然语言处理等技术进行替代，可以高效地抽取非结构化数据中的高价值信息，大大减少重复繁琐的信息收集工作，进一步提升金融从业者的工作效率。

3.3 市场预测应用

机器学习在市场预测领域的应用，一般采用与专家经验相结合的方式进行建模，其应用的核心是对指标和价格走势等的预测。其中，模型输入将市场各种可量化数据进行了深度融合，输出则给出市场指标和价格的预测走势，进而辅助量化交易或其他需要依赖金融市场的决策。进行市场预测的目的旨在减少未来金融活动的不确定性，降低决策可能遇到的风险，提高经营与管理的科学水平，减少决策的盲目性，使决策目标得以顺利实现。

4 智能运营

4.1 单证识别应用

传统金融机构在资金结算、贸易融资、保险理赔等业务场景当中，需要将转账支票、现金支票、信用证、保险单等大量凭证信息录入系统。传统业务办理流程中多为人工录入信息，人力耗费量大、识别率低、准确率低且录入时间长。应用图像识别技术，可实现票据识别、证照识别和通用印刷体、手写体的文字识别。目前使用图像识别技术进行凭证识别可大量代替人工，并且提高识别率和准确率，极大地节约了人力成本，提升了运营效率。

4.2 银行网点实体服务机器人应用

在银行网点或无人银行里，机器人大堂经理可以识别客户身份，通过语音交互为客户预约办理业务，并引导客户进入不同服务区域。目前银行网点服务机器人可实现智能分流取号、引导下载手机银行和使用自助设备、银行业务预处理等功能。银行网点服务机器人会主动询问顾客业务需求，引导顾客使用银行网点自助设备，进而促进网点离柜业务的高效转化，降低实体网点运营成本。机器人亲和友好的外形可引发客户的关注和好奇，加上主动与顾客交流的特质，赋予了它良好的营销能力，可提高银行网点金融产品的营销成功率。

4.3 自动话务质检应用

在话务质检和分析场景，目前主要使用人工听录音的方式进行审核，存在成本高、效率低和关键信息错失的问题，运用智能语音技术和数据分析技术，可将坐席和客户的交互语音结构化，对坐席的服务行为和合规性进行自动化质检，并挖掘分析有价值的信息，为服务和营销等提供数据决策支持。

4.4 实时坐席辅助应用

在坐席和客户交互场景，坐席解答客户问题时，一般根据经验和可搜索型知识库，存在对外服务不统一、服务效率低和无法实时监控坐席服务行为的问题。运用智能语音技术与自然语言处理技术，可实时理解客户意图，从知识库中获取答案推送到坐席端，并且可实时监控坐席的服务行为，达到降低坐席人员的培训、管理成本，并提升客户服务体验的目的。

4.5 车险产品定价应用

保险公司可通过客户信息、驾驶行为信息、车辆定位轨迹信息，预测驾驶速度，并结合其他的非保险领域的数据，建立定价模型，自动分析其风险系数，给出更加精准的定价。在车险市场竞争日趋激烈的环境下，基于机器学习的差别化、精准的定价将发挥越来越大的作用。

4.6 保险车辆定损应用

保险公司可通过人脸识别技术、语音识别技术识别客户身份，分析车辆损坏部位照片，并进行车损部位高精度图片识别、秒级定损、自动精准定价、智能风险拦截。同时，采用机器学习辅助车辆定损，可减少查勘定损人员大量工作，大幅度提高定损效率、缩短赔付时间，提升客户体验，并且可有效辅助识别欺诈案件。

5 监管与合规

5.1 银行反洗钱应用

面对跨境网络支付、贸易融资等虚构交易伪造手段，反洗钱监管压力日益增大。长期以来，国内金融机构主要针对大额可疑交易行为进行反洗钱监管，但缺乏足够的历史交易数据对这些可疑账户的洗钱行为趋势、资金的往来路径做出预测和风险预警。

基于机器学习的反洗钱监测模型，通过多维度监测可疑交易，统筹考虑相关人物、事件、内容、地点、时间、原因等要素，分析交易行为的特点，可协助银行有效识别可疑交易，完善反洗钱风险管理体系。

5.2 智能合规审计应用

合规审核方面，可通过自然语言处理等技术将监管规则数字化以方便机器执行，并实时跟踪对比海内外不同市场监管规则的异同及更新，同时基于生物识别及知识图谱等技术提高客户身份识别效率和质量，降低合规审核成本。合规分析方面，可基于智能推理及预测模型等提示机构操作合规风险并优化决策过程。

参 考 文 献

[1] GB/T 5271.31—2006 信息技术 词汇 第31部分：人工智能 机器学习

[2] GB/T 22081—2016 信息技术 安全技术 信息安全控制实践指南

[3] JR/T 0221—2021 人工智能算法金融应用评价规范

[4] ISO/IEC 27033 信息技术 安全技术 网络安全（Information technology—Security techniques—Network security）