

来源：2022第三届中小金融机构数智化转型优秀案例评选

获奖单位：张家港农商银行

荣获奖项：数据智能应用创新优秀案例奖

一、项目方案

推动数智化转型是我行把握数字经济、数字金融，迈向高质量发展的重要抓手，而充分挖掘并应用数据资产的价值，则是数智化转型的重要基石。征信信息作为我行在经营过程中积累的重要数据资产，基于二代征信数据建设衍生指标库是我行锤炼数字化竞争力，发展金融科技的核心工作之一。

通过对征信数据的挖掘、加工和衍生相关指标，全面提升我行征信数据应用标准化、敏捷化、精细化、安全化的程度。利用衍生指标体系，精准刻画客户画像、快速迭代风控规则、有效匹配客户需求。项目的落地提升了我行对风控模型、策略等的迭代效率与识别能力，实现了征信数据的闭环管理，在安全合规的基础上充分挖掘数据资产的价值，是我行数智化转型的重要举措。

1.项目建设整体目标

本项目的建设整体目标：基于二代个人征信数据，设计征信指标衍生变量，形成一套完善、稳定、具有强风险区分能力、强业务可解释性的征信特征指标体系，充分挖掘征信信息，提高风险计量水平和风险管理能力。

2.项目建设主要任务

（一）支撑一二代征信准确切换

项目的落地为我行快速、准确地实现一代二代征信报告应用规则和风险模型切换提供了重要保障。通过前期对二代征信结构表和指标的梳理，有效支撑一代征信指标到二代征信指标的快速映射。

（二）支撑业务多样化风控需求

二代征信相较于一代征信展示内容明显增加，本项目实现了对二代征信报告信息的充分挖掘，能够覆盖业务所需的各项征信规则指标，为业务全流程模型及规则风控需求提供便捷。通过本项目，实现我行各业务条线产品二代征信特有风控指标、风控规则的落地。

（三）提升风控模型及策略迭代效率

时点指标使用逻辑的一致性，极大地优化了模型策略迭代的延续性和效率。

（四）为后续征信数据挖掘与分析奠定基础

本项目一期搭建了征信指标库、二期进一步构建了客户征信标签库，进一步简化、脱敏征信数据，一方面有助于我行对资产客户进行分层分群，全面了解资产客户的风险画像；另一方面也有助于我行开展风控或营销，把控客户风险、提升客户粘性，降低客户营销成本。

3.项目架构

（一）指标库架构设计

整体解决方案采用了“三表一体系”的架构设计，即原始表、基础表、中间表、征信衍生指标体系。该架构设计保证了指标体系的可拓展性、不同业务的适用性、整体系统的灵活性。基础表的设计兼容一代、二代原始表结构，保证一代、二代映射关系的一致性和准确性；中间表拥有完整的SDM，只需处理基础表即可快速落地；指标衍生体系在中间表的设计基础上进行衍生,最大限度地从原始数据中提取特征，保证指标体系的可扩展性。

该架构保证了指标全面性的同时，兼顾了体系的易扩展性与灵活性，满足了对业务部门的敏捷化支持，当业务提出新的指标需求，基础表与中间表的加入可极大简化

开发流程，只需将加工逻辑加入衍生体系即可做到指标的快速落地与复用。

（二）指标维度体系设计

项目设计了人行征信变量衍生体系，维度丰富且覆盖全面，有利于挖掘显著风险区分能力的指标，支持银行灵活配置和管理衍生变量。本项目基于对征信数据的充分挖掘，包含个人信息类、账户行为类、逾期类等12个指标大类，细分为37个指标小类，形成衍生指标1600多个，实现对征信信息精细且全面的覆盖。

二、项目创新点

1.指标体系的完备性

本项目在构建过程中综合考量人行征信报文结构和指标的业务应用两个维度，确保能够充分挖掘人行征信的信息，提升数据的信息抽象能力。

2.特征构建的全面性

本项目分别从业务品种/账户类型、
开立时长、统计指标

（逾期、查询、额度、账户等）、统计维度（时间、机构类型、担保方式、账户状态等）、聚合方式等维度进行笛卡尔积式特征构建，以保证特征构建的全面性。

3.特征变量的衍生性

通过特征组合、交叉等方式，生成如：申得率、使用率、申得差值等组合特征，组合特征能够多维度地评判风险，进一步进行特征衍生，提升特征的区分能力。

4.指标逻辑的统一性

扎口全面层面的业务场景接入，保证不同部门、不同团队使用统一的指标逻辑，极大提升了风控效率；

5.指标扩展的灵活性。

衍生体系的架构保证了新增指标的快速落地与复用。

6.底层数据平台的自主性

本项目充分运用大数据处理能力建立先进的数据处理平台，基于完善的国产化Gauss数据库，构建了全行层面的征信数据应用解决方案，稳定可靠、节约成本、响应快速，为我行掌握独立自主的信贷风控能力奠定了良好的技术基础。

三、技术实现特点

1.关键技术的高效性

（一）灵活的微服务架构

采用微服务架构的方式，实现可插拔式布局，并能与现有系统实现轻耦合对接。产品前端基于微服务架构设计，借助诸如BI平台等行内现有服务架构，实现基于WEB的指标管理及应用。

（二）稳定且易复用的流式计算处理技术

采用流式计算处理技术，逻辑清晰易复用，极大提高了运行稳定性及性能。后端基于国产化Gauss数据库实现对源表的流式计算处理，对全部特征指标的算法逻辑进行抽象，共提炼了7种不同类型的指标算法，并进行分布式hash计算存储管理，最大程度降低了对数据库算力的依赖。

2.架构设计的前瞻性

产品充分考虑到征信体系未来发展的需要，采用分布式指标处理框架，实现未来指标扩充的快速便捷。架构中后台基础表、中间表、指标衍生体系都可进行横向扩展，能根据实际情况对指标进行快速升级、扩容以达成对不同业务场景的敏捷化支持。同事得益于流式数据处理技术，指标加工流程代码相对独立，未来增、减后台表无需修改现有代码。

3.体系设计的统一性

项目通过全行级的指标体系统一管理，实现了不同业务条线间指标逻辑的统一。相较于阅读传统征信源报告，标准化的指标体系使业务人员能够更为直观且便利地提取征信数据中高价值部分。

4.数据应用的安全性

针对征信数据敏感性的需求，一方面对敏感信息进行脱敏处理，另一方面项目采用完全隔离的方案模式。前端可针对的业务进行差异化权限配置，后端整体流程只对行内技术人员开放。运行过程中所涉及所有数据及信息只会留存在内网中，与外部网络完全隔离，协助管理者对指标增删改查的安全管理，确保不会泄露任何客户信息。

四、项目过程管理

张家港农商银行二代征信衍生指标库项目整体实施周期7个月，项目于2021年08月正式启动，分别在2021年11月完成三表架构的上线，2022年2月完成征信衍生指标体系的上线。项目整体可划分为7个阶段，具体如下：

1.项目启动

项目启动阶段1-2周。主要工作内容为组搭建项目推进组织架构、明确项目牵头部门和配合部门、成立项目推进小组、制定项目计划、建立沟通及汇报机制等。

2.需求分析

需求分析阶段4周。主要工作内容为搜集业务条线需求、梳理行内现行的风控模型及风控规则涉及的指标、制定需求文档等。

3.架构设计

架构设计阶段6周。主要工作内容为架构设计，明确征信源表、基础表、中间表映射关系及指标衍生体系加工逻辑，产出相关设计图表，进行架构设计评审等。

4.项目开发

项目开发阶段3个月。主要工作内容为针对项目开发内容进行分工，并依照需求内容分配开发任务、编写开发文档，依照项目架构编写指标加工代码等。

5.系统测试

系统测试阶段4周。主要工作内容为编写测试用例、核对指标数据、测试过程性能及调度稳定性、修改测试问题并进行代码调优、产出测试报告等。

6.上线投产及试运行

系统试运行阶段1-2月。主要工作内容为整理上线文档、上线投产、回输历史数据、技术及业务验证、系统推广等。

7.系统运维

主要工作内容为运维监控系统运行情况并记录。

五、运营情况

1.系统运行情况

该项目自2021年8月正式启动以来，项目进度有序推进。在数据采集、过程开发、联调测试、业务诊断咨询、规则策略制定、流程架构及体系规划、性能调优等方面进行过多次论证，为行内对于征信数据采集技术、数据处理及挖掘技术、特征工程技术奠定了坚实基础。

2.推广应用

征信特征指标现作为核心数据资产的重要组成部分，已实现了多项后续推广应用。

（一）客户画像分层分群

标签库

通过对征信指标信息进一步的分类分级，将行为化的征信指标数据加工为特征化的标签数据，刻画客户行为画像，对客户分层分群，实现标签化管理。相较于传统信贷审批、贷后管理等征信应用，标签体系可简化征信信息，但同时保留重要的客户行为特征，有效应用于精准营销、客户分群等层面，为行内产品营销、客户信贷需求转化等工作提供支持。

（二）信贷产品指标分库建设

通过对征信指标数据的进一步整合，我行现已针对行内大部分信贷产品分别构建其准入、授信、用信、贷后管理征信规则及评分卡指标分库。将整体的一级指标库进一步细化为针对各信贷产品的二级指标库，一方面可以高效率地对行内征信进件进行全产品维度的预跑批，最大程度挖掘征信进件价值，提升行内产品营销、客户信

贷需求转化的成功率；二能实现对贷后征信的自动化预警，大大提高了我行风险监控精度，提升了我行风险防范能力。

六、项目成效

征信指标体系已成为我行信用风险管理的重要基础，极大地加强了我行信用风险信息收集、分析及转化应用水平。

1.支持全行信贷客户分群

通过贷前审批、贷后管理等流程，加工客户征信指标和标签信息，实现对资产客户的分层分群。目前，我行已根据客户贷款情况、信用卡情况、征信查询情况、我行信贷产品情况等将客户划分为162类客群，有效支撑我行客户的精细化管理。

2.支持全行风控模型的快速开发

征信特征体系是我行风险数据集市的重要组成部分，指标体系上线后，省去了前期征信数据清洗加工的流程，极大提高了评分卡开发的效率。征信指标上线后，我行已累计完成了11个模型的开发上线，模型开发周期由之前的平均3个月提升到目前的1个月，模型开发效率提升了3倍。

3.支持线上信贷业务贷前风险策略高效迭代

我行线上信贷业务当前处于高速发展阶段，更大的业务范围和更广的产品受众对贷前风控提出了更为严格的要求。得益于标准化的设计框架，业务可针对性地获取所需的征信指标数据，避免二次加工，极大提升了策略迭代的效率。开发人员可通过清晰的逻辑映射进一步简化并缩短新策略的加工及上线流程。征信指标上线后，我行已累计对网金及小微条线部分产品贷前策略进行过多次优化和迭代。

4.完善贷后预警策略设计

征信指标上线后，我行有针对性地将征信指标与现有的行内客户信贷信息、行内信贷贷后行为信息等指标相互融合，完善了我行的贷后预警策略。相较于传统的查阅征信报文，标准化指标的加入可更为客观全面地评价客户贷后风险，并对突出客户进行自动化预警。现贷后预警策略已优先设计应用于我行自营互联网金融产品。

5.充分挖掘客户信贷需求

通过信贷产品征信指标分库的建设，我行逐步实现全产品的征信规则测算。目前，征信指标分库已覆盖12个产品，摆脱客户进件渠道的束缚、打破行内业务条线的壁垒，最大化匹配客户的信贷需求、最大化提升客户的实际体验、最大化节约客户的营销成本。

七、经验总结

伴随着近年来金融科技的飞速发展，算法、算力的不断提升，金融数值化转型已然成为大势所趋。而“数据驱动”则是检验数智化转型深度与质量的试金石，对于数据价值的深层次挖掘与安全高效的应用则是实现“数据驱动”的核心基础。

作为数智化转型的奠基石，针对智能化数据应用的系统设计不仅应具备标准化、易用化、灵活化的特征，更应当在合规安全的基础上保证系统的高效性、可靠性、扩展性、前瞻性。伴随着金融技术的不断完善革新，数据挖掘技术的不断发展，未来的征信指标库的应用范围会更加丰富，

打破数据孤岛

、突破渠道和条线限制，成为我行精准营销、有效风控以及高效服务三大领域的重要基石。

更多金融科技案例和金融

数据智能优秀解决方案，请登录[数字金融创新知识服务平台-金科创新社](#)官网案例库、选型库查看。