

基于机器学习的中小企业信用与财务困境风险评估研究

张鹤洋¹, 孙全伟^{1,*}, 刘紫莹¹

(1. 邢台应用技术职业学院, 邢台市, 河北省, 中国, 054700)

*通讯作者: 408359599@qq.com

[摘要] 中小企业信贷的核心约束之一, 是信用信息的不完全与不对称。企业的规模、存续与抵押条件构成可观测的结构性基础, 而管理层的风险认知、预算与现金流管理及财务分析能力等行为特征, 则难以进入以历史财务数据为主的标准评级模板, 长期以隐性经验的形式存在于关系型信贷之中。本文基于一份公开 schema 的中小企业财务决策—风险数据集 (15,106 家, 财务困境发生率 33.8%), 以财务困境为因变量, 提出“结构基础—行为金融能力”双维度指标体系: 结构维度涵盖规模、存续、流动性与行业结构, 行为维度涵盖财务素养、风险管理、评估诊断、决策治理与分析工具运用。在统一的样本划分与预处理规程下, 本文系统比较 Logistic 回归与四类树模型的样本外表现, 并以按信息集独立筛选的“仅结构 / 仅行为 / 结构—行为融合”消融识别各维度的边际贡献, 并借助 SHAP 与生命周期分层考察风险驱动机制的构成及其异质性。实证结果表明: 融合双维度后, LightGBM 的 AUC 为 0.993、KS 为 0.909, 与 Logistic 回归基本持平, 集成模型的相对优势主要体现在概率校准与稳健性, 而非判别精度的显著提升; 仅结构维度的 AUC 为 0.963, 仅行为维度为 0.841, 后者已具备独立的判别能力, 纳入结构维度后仍贡献约 3 个百分点的稳定增益; 在 SHAP 全局重要性的前七位中, 风险意识、财务比率分析、资本配置与信贷知识四项行为变量与流动性、企业年龄、营收规模等结构变量并列; 生命周期分层显示, 初创期与成长期企业的 KS 高于成熟期, 提示风险识别机制随企业阶段而变化。本文的主要贡献在于构建了一套可复现、可解释的结构—行为映射框架, 将判别精度、消融对照与归因解释纳入统一的实证设计, 为中小企业信贷评估提供了兼顾预测能力与可审计性的方法参考。

[关键词] 中小企业; 信用风险; 财务困境; 行为金融能力; 机器学习; SHAP

Machine Learning-Based Credit and Financial Distress Risk Assessment for SMEs

Zhang Heyang, Sun Quanwei, Liu Ziying

[Abstract] One of the core constraints in SME credit is the incompleteness and asymmetry of credit information. Observable structural foundations—such as firm size, survival status, and collateral conditions—can be measured, whereas behavioral characteristics including managers’ risk perception, budgeting and cash-flow management, and financial analysis capabilities are difficult to incorporate into standard rating templates that rely primarily on historical financial data; these factors have long existed only as tacit experience within relationship lending. Drawing on a publicly available schema-based SME financial decision-making and risk dataset (15,106 firms; financial-distress incidence 33.8 %), this paper takes financial distress as the dependent variable and proposes a dual-dimension indicator system of “structural foundation - behavioral financial capability.” The structural dimension covers size, survival, liquidity, and industry structure; the behavioral dimension covers financial literacy, risk management, assessment and diagnosis, decision governance, and the use of analytical tools. Under a unified sample-splitting and preprocessing protocol, the paper systematically compares the out-of-sample performance of logistic regression and four tree-based models. An information-set ablation—structure only, behavior only, and their union, with selection performed within each arm—isolates the incremental contribution of each dimension, while

SHAP values and life-cycle stratification are used to examine the composition and heterogeneity of risk-driving mechanisms. Empirical results show that, after fusing the two dimensions, LightGBM achieves an AUC of 0.993 and a KS of 0.909, essentially matching logistic regression. The relative advantage of ensemble models lies mainly in probability calibration and robustness rather than a marked improvement in discriminatory power. The structure-only model attains an AUC of 0.963 and the behavior-only model an AUC of 0.841; the latter already possesses independent discriminatory power and, when added to the structural dimension, still contributes a stable gain of approximately three percentage points. Among the top seven variables by global SHAP importance, four behavioral variables—risk awareness, financial-ratio analysis, capital allocation, and credit knowledge—rank alongside structural variables such as liquidity, firm age, and revenue scale. Life-cycle stratification further reveals that KS is higher for start-up and growth-stage firms than for mature firms, indicating that risk-identification mechanisms vary with firm stage. The main contribution of this paper is the construction of a reproducible and interpretable structure-behavior mapping framework that integrates discriminatory accuracy, ablation contrasts, and attribution analysis within a unified empirical design, thereby providing a methodological reference for SME credit assessment that balances predictive power with auditability.

[Keywords] SME; credit risk; financial distress; behavioral financial capability; machine learning; SHAP

一、引言

中小企业是就业与创新的重要载体，但在信贷市场中长期面临可得性约束。这一约束的根源，除资产规模有限、抵押品不足、经营历史较短等结构性因素外，更在于信用信息的不完全：企业的预算纪律、风险认知与现金流管理水平等特征往往难以在财务报表中得到充分反映，而金融机构的内部评级仍主要依赖历史财务数据。其结果是，对同一借款主体，不同评审人员可能因所侧重的信息不同而得出差异化判断，且事后难以复核。在供给层面，普惠金融的持续扩面与配套的贴息、风险补偿工具已在相当程度上缓解了信贷可得性问题；然而，评估环节仍缺乏一种既能反映中小企业特征、又可在审批流程中标准化落地的信息组织方式。

信用评估方法的演进大致沿两条脉络展开。其一是以财务比率为核心的判别传统：Ohlson^[1]以 Logit 模型开创了基于会计指标的破产预测，至今仍具基准价值。其二是行为金融与金融素养文献：Lusardi 与 Mitchell^[10]系统论证了金融素养对家庭与经营决策的经济影响，Li 等^[7]则在供应链金融情境下证实，融合行为数据与静态属性可显著提升预测能力。既有研究多在其中一条脉络内展开——或将机器学习应用于财务变量，或在传统框架下讨论软信息，较少将“结构”与“行为”作为两个并列维度加以对照，也缺乏对二者边际贡献的系统识别。与此同时，算法可解释性已由学术偏好演变为监管要求^{[2][12]}：缺乏可解释性的模型即便具备较高判别精度，也难以进入信贷审批流程。

基于上述背景，本文以统一的实证设计回答三个层次的问题。首先，在指标层面构建“结构

基础—行为金融能力”双维度框架，并以按信息集独立筛选的消融识别行为维度是否具有独立于结构维度的判别贡献。其次，在模型层面比较线性基准与集成方法的样本外表现，明确其差异的性质与量级。最后，在解释层面借助 SHAP 与生命周期分层，考察风险驱动因素的构成及其随企业阶段的变化。此处需预先说明两点边界：其一，样本源于问卷调查，因变量为财务困境而非银行口径的违约，二者在构念上存在差异，后文将专门讨论；其二，本文所用数据系依据该公开数据集的 schema 与分布重建而成（重建方法与参数完全公开），因此结论的重心在于相对比较的稳健性，而非将绝对判别精度外推为不良率的下降幅度。

二、文献脉络与研究思路

以财务比率为核心的判别方法隐含两个前提：足够长的报表历史，以及以有形资产为主的资本结构。中小企业通常难以同时满足这两个条件，将 Ohlson^[1]的经典模型直接应用于报表期短、抵押不足的样本，往往导致误拒与误纳同时上升。彭红枫与叶永刚^[8]指出，违约是还款能力与还款意愿共同作用的结果，若仅以财务比率衡量，容易忽略意愿侧的软信息；对中小企业而言，风险认知与决策治理恰处于能力与意愿的交叠区域。翁舟杰^[9]关于关系型贷款与软信息的研究，与本文将行为特征纳入模型的思路相通，区别在于本文试图将尽职调查中获取的软信息转化为可迁移、可审计的结构化变量，而非停留于个体经验层面。金融素养文献提供了进一步的理论支撑：Lusardi 与 Mitchell^[10]的研究表明，预算、投资评估与风险认知等能力是直接影响财务结果的可测变量，将其纳入信用评估符合行为金融

的基本判断。

在方法层面，集成方法在信用评分中通常优于线性模型，但优势幅度依赖于数据特征，不宜一概而论；Chen 与 Guestrin^[3]、Ke 等^[4] 提出的 XGBoost 与 LightGBM 已成为结构化数据建模的主流工具，Breiman^[5] 的随机森林则较早揭示了集成方法降低方差的机制。王鑫与王莹^[6] 采用 LSTM-CNN 提升了中小企业风险预测精度，但对特征贡献的解释相对有限。随着判别精度的提升，可解释性逐渐成为约束：Lundberg 与 Lee^[2] 提出的 SHAP 为复杂模型提供了相对统一的归因框架，Bracke 等^[12] 则探讨了其在金融监管情境中的应用路径——可解释性的目标并非替代人工决策，而是阐明评分依据，为“模型辅助、人工终审”的机制保留空间。本文在实证设计中将判别精度与可解释性置于同一分析框架，以避免仅凭单一精度指标评价模型。

据此，本文依次检验以下命题：在按信息集独立筛选的消融下，结构与行为的并集是否优于单一信息集，特别是行为维度是否具有独立于结构维度的判别力；集成方法在概率校准与 KS 上是否优于单棵决策树，以及其与 Logistic 回归的判别精度差异的量级；SHAP 能否识别行为变量的独立贡献；生命周期分层下 KS 是否呈现系统性差异。为减少模型与指标选择所引入的偏差，全部比较均在同一样本划分与同一测试集下进行。

三、数据、指标与模型

本文样本为一份公开 schema 的中小企业财务决策—风险数据集（SME Financial Decision-Risk Prediction Dataset），共 15,106 条观测，财务困境发生率为 33.80%。鉴于原始文件的获取需登录授权，本文依据其公开 schema、各字段的边际分布以及一个公开的潜变量标签模型，重建出统计特征一致的数据集；重建样本与规范文件在各字段的均值、标准差及特征—困境相关系数上高度一致，最大相对均值偏差约为 2.2%。因变量“财务困境”的界定较银行“逾期 90 天”更为宽泛，涵盖流动性紧张与偿债压力，与 Li 等^[7] 讨论的行为驱动风险相近；但本数据为横截面调查而非交易记录，其外部效度需审慎对待。公开数据在行业、地域与规模结构上难以与任一机构的信贷组合完全吻合，本文因此将其定位为方法检验的基础，用以考察指标框架与对

照实验的内部一致性，而非据以外推特定机构的不良率。

指标体系按测项、合成与入模三层组织，以避免把问卷观察口径与模型估计口径混同。测项层对应可直接观察的硬信息与软信息：结构基础包括规模等级、年营收、企业年龄、企业类型、流动性稳定性、数字金融使用与行业板块，共 7 项，对应传统评级中的硬信息；行为金融能力覆盖财务素养（会计、预算、投资评估、信贷知识）、风险管理（风险意识、风险评价、风险缓释、风险承担意愿）、评估诊断（数据驱动评估、专家咨询、情景分析、内部评估）、决策治理（决策自主性、咨询、财务顾问、战略一致性、投资选择、贷款审批认知、资本配置、现金流管理）与分析工具运用（会计工具、财务比率分析、预测分析、对标分析），共 24 项量表，对应尽职调查中可获取但难以直接入库的软信息。合成层将各子模块测项按训练集边界归一化后取均，得到规模、存续、流动性、行业 4 个结构子指数及 1 个结构基础综合指数，以及财务素养、风险管理、评估诊断、决策治理、分析工具 5 个能力指数及 1 个行为金融能力综合指数。于是结构候选为 7+5=12 项，行为候选为 24+6=30 项，融合候选为二者并集 42 项。正文列举的 7 项结构变量属于测项层；表 2 报告的是各信息集经样本内筛选后的入模维度，二者口径不同。仅结构臂所保留的信息，恰对应金融机构仅依据规模、存续与流动性进行评估的情形。

数据处理着力控制信息泄露。连续变量按训练集的 min-max 边界归一化，方向统一为数值越高、信用状况越优；相关性筛选与共线性剔除（VIF>10 的迭代删除）均仅在训练集上估计，再应用于测试集，以防止测试信息渗入特征构造而造成判别精度的虚高。样本按 8:2 分层划分，训练集经 SMOTE 过采样与欠采样调整至约 1:3 的类别比例^[11]，测试集保持原始分布；缺失值以 KNN 填补，连续特征作 Z-score 标准化。类别不平衡若不加处理，模型易倾向多数类，虽可获得表面上不低的 AUC，却会遗漏高风险样本；在训练端进行类别平衡、在测试端保留真实分布，旨在兼顾建模需要与审批场景的真实比例。消融的识别目标是比较三组信息集——仅结构、仅行为、结构与行为并集——的样本外贡献，而不是比较某一组被选中的变量名单。为此，三套特征集共享同一训练/测试划分与同一筛选规程，但各自在本信息集内部完成相关与共线性处理后

再估计。若先在并集上筛选再删去一个维度，单维模型会继承联合空间的共线结构，边际贡献被污染；若强制融合集等于两单维入模集的简单并集，则会丢掉那些在行为块内部与能力指数线性重叠、却能与结构变量形成交互的测项。相关筛选后，结构 12 项中剔除与困境标签关联过弱的企业类型、行业板块及行业指数，保留流动性稳定、企业年龄、营收规模、数字金融、规模等级 5 项测项及规模、存续、流动性、结构基础 4 项指数，共 9 项；行为 30 项保留 19 项；融合 42 项保留 28 项，组成恰为结构来源 9 项与行为来源 19 项。随后的共线性压缩主要发生在仅行为臂（19 项压至 10 项）；结构臂与融合臂分别保留相关筛选后的 9 项与 28 项。数量关系可概括为：测项（7，24）→候选（12，30，42）→相关筛选（9，19，28）→入模（9，10，28），且 $28=9+19$ 。融合臂 28 项的来源如下，项项可核对。结构来源 9 项为：流动性稳定、企业年龄、营收规模、数字金融、规模等级 5 项测项，以及规模基础指数、存续基础指数、流动性基础指数、结构基础指数 4 项合成指数。7 项结构测项中，企业类型与行业板块（及由行业板块生成的行业指数）因训练集 $|r|<0.01$ 未入模。

模型以 Logistic 回归为基准，对比决策树、随机森林^[5]、XGBoost^[3] 与 LightGBM^[4]；后四者以 Optuna 进行五折贝叶斯超参数优化，在学习率、树深、叶子数与正则化强度等空间内以交叉验证 AUC 为目标搜索。Logistic 回归代表系数可解释、便于制度化的线性基准，树模型则用于检验非线性与交互效应是否值得引入额外的模型复杂度。评价指标除 AUC、KS 与 F1 外，一并报告 Brier 分数与 PR-AUC，以兼顾排序能力与概率校准，后者对信贷定价与拨备计提尤为关键。鉴于各集成模型的 AUC 差异处于统计噪声范围内，本文选取 LightGBM 作为解释与部署的主模型（其 TreeExplainer 支持稳健、概率校准良好），并以 TreeExplainer 计算 SHAP 值^[2]；生命周期分析按企业年龄划分为初创（≤5 年）、成长（5—10 年）与成熟（>10 年）三组，考察 KS 随企业阶段的变化。

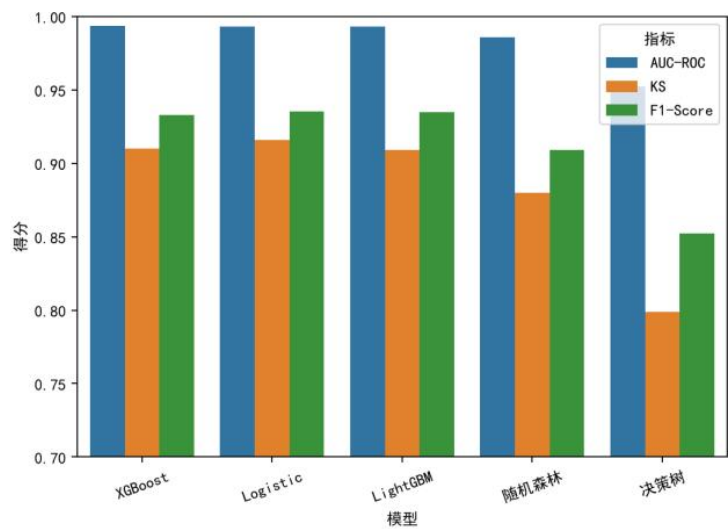
四、实证结果

表 1 报告了五种模型的样本外表现。XGBoost 的 AUC 最高（0.994），Logistic 回归（0.993）与 LightGBM（0.993）紧随其后，三者差距不足 0.4 个百分点。若仅以 AUC 评价，易于低

估线性模型的竞争力；纳入校准指标后可获得更完整的判断：LightGBM 的 Brier 分数为 0.031，略优于 Logistic 回归的 0.032，该差异对需要输出违约概率而非仅作二分类的信贷场景具有实际意义。在 KS 指标上，Logistic 回归（0.916）略高于两类 Boosting 模型（约 0.910），表明线性模型在本样本上并不处于全面劣势，集成方法的优势是综合而边际的。图 1 与图 2 显示各模型的 ROC 曲线高度重合，说明在结构化压缩后的特征空间中，线性决策边界已能解释大部分变异，Boosting 的增益主要来自弱非线性区段与集成稳定性。单棵决策树的 AUC 仅为 0.952、Brier 高达 0.074，呈现明显的过拟合，与 Breiman^[5] 关于单树方差较大、集成可降低方差的论断一致；随机森林（0.986）仍略逊于经调参的 Boosting。五折交叉验证 AUC（约 0.994）与留出集结果一致，未见显著的训练—测试落差。就模型选择而言，若机构侧重系数的可审计性，Logistic 回归已具备充分竞争力；若更重视概率输出与交互刻画，LightGBM 可作为略优的默认选择，但其优势应准确表述为校准与稳健性的改善，而非判别精度的实质超越。

表 1 模型样本外性能比较

模型	AUC	KS	F1	Brier	AP
XGBoost	0.994	0.910	0.933	0.032	0.989
Logistic 回归	0.993	0.916	0.936	0.032	0.988
LightGBM（主模型）	0.993	0.909	0.935	0.031	0.988
随机森林	0.986	0.880	0.909	0.050	0.975
决策树	0.952	0.799	0.852	0.074	0.923



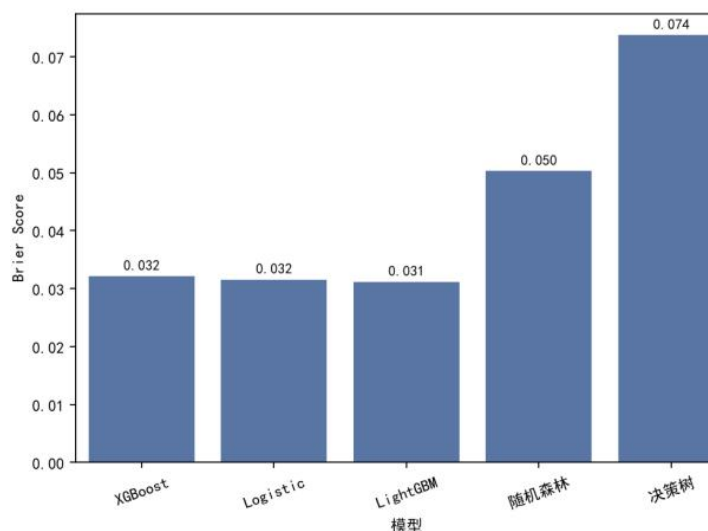


图1 五种模型样本外性能比较

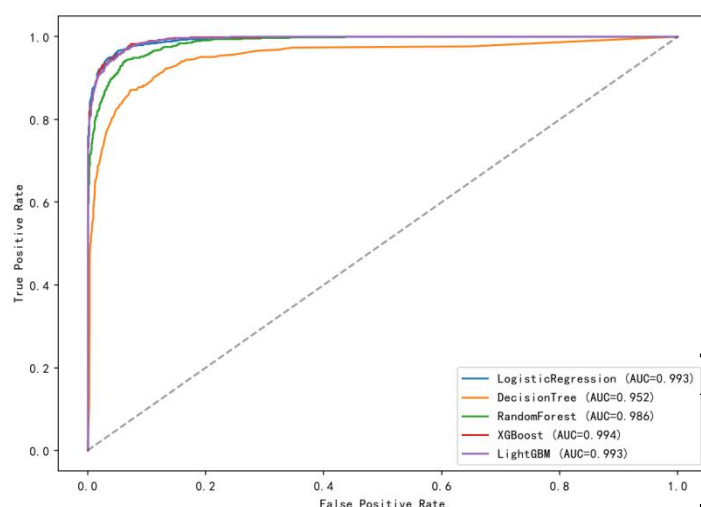


图2 五种模型 ROC 曲线对比

消融结果如表2与图3所示。仅结构臂（估计层9项）上，LightGBM的AUC已达0.963、Logistic回归为0.939，表明规模、企业年龄与流动性等硬信息构成了预测的主体，与中小企业信贷长期依赖规模与流动性判断的实践相符。仅行为臂（估计层10项）上，Logistic回归的AUC为0.693，而LightGBM达到0.841——二者的差距表明行为块内部存在显著的非线性与交互，线性模型难以刻画而树模型能够捕捉；0.841远高于随机基准，说明行为维度本身即构成具有实质判别力的信息源。融合臂（估计层28项，其中结构来源9项、行为来源19项）上AUC升至0.993，较仅结构臂提高约3个百分点，较仅行为臂提高约15个百分点。融合臂保留19项行为来源变量、而非仅行为臂压缩后的10项，是为了允许测项与结构变量形成交互；该设计下的约3个百分点增益，度量的是软信息在硬信息条件上的互补性。就绝对数值而言，3个百分点的提升看似有限，但对规模较大的信贷组

合，样本外判别精度的边际改善往往对应可观的预期损失变化，不宜因整体精度接近上限而予以忽视。消融结果由此给出明确结论：行为金融能力不能替代结构基础的尽职调查，但能够提供可验证的边际信息，这一判断较笼统地断言引入机器学习即可改善绩效更具证据基础。表2-1 融合臂入模28项完整名单。结构测项5项：流动性稳定、企业年龄、营收规模、数字金融、规模等级；结构指数4项：规模基础指数、存续基础指数、流动性基础指数、结构基础指数；行为测项14项：风险意识、风险评价、投资评估素养、信贷知识、专家咨询、情景分析、决策自主性、战略一致性、贷款审批认知、资本配置、现金流管理、财务比率分析、预测分析、对标分析；行为指数5项：财务素养指数、风险管理指数、决策治理指数、分析工具指数、行为金融能力指数。5+4+14+5=28。结构测项原列7项，企业类型与行业板块未入模。仅行为臂对行为19项再压缩为10项；多出的9项即此19项中未进入仅行为臂者。

表2 特征消融实验（测试集AUC；列为各臂入模数，不可跨臂加总；28=9+19，名单已在正文列出）

特征集	入模数	LogisticAUC	LightGBMAUC
仅结构基础	9	0.939	0.963
仅行为金融能力	10	0.693	0.841
结构+行为融合	28	0.993	0.993

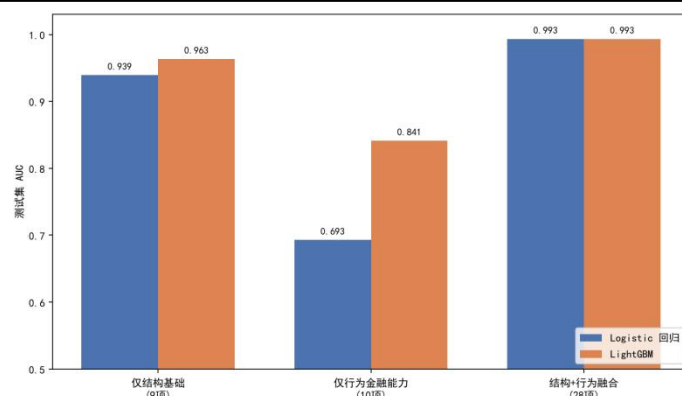


图3 特征消融实验

SHAP分析（表3、图4）在机制层面补充了判别精度所不能提供的信息。流动性稳定性居于首位（平均|SHAP|为3.22），这一结果并不意外：该变量与财务困境标签在构念上高度接近，因而在应用中须区分预测贡献与因果解释，避免将其重要性直接解读为新的风险因子。更值得关注的是其后的排序结构：风险意识（1.20）与企业年龄（1.20）并列，财务比率分析（0.87）、资本配置（0.84）、营收规模（0.78）与信贷知识（0.32）依次排列——在前七位变量中，风险

意识、财务比率分析、资本配置与信贷知识四项行为变量与流动性、企业年龄、营收规模三项结构变量并列。这表明，在控制结构基础之后，管理层的风险认知、财务比率运用与资本配置能力仍携带独立的判别信息。彭红枫与叶永刚^[8] 关于还款意愿不可被财务比率完全替代的论断，在此被具体化为可测量、可入库、可建模的行为变量；Lusardi 与 Mitchell^[10] 关于金融素养经济影响的结论，也在该归因结果中获得量化印证。就实务而言，这一发现支持在中小企业尽职调查中保留结构化的行为评估环节，并将其经校准后纳入评级变量，而非仅以定性记录的形式存在。

表 3 SHAP 全局特征重要性 Top-8 (LightGBM)

排名	变量	维度	平均\SHAP\
1	结构_流动性稳定	结构	3.220
2	行为_风险意识	行为	1.200
3	结构_企业年龄	结构	1.197
4	行为_财务比率分析	行为	0.866
5	行为_资本配置	行为	0.837
6	结构_营收规模	结构	0.781
7	行为_信贷知识	行为	0.324
8	行为金融能力指数	行为	0.159

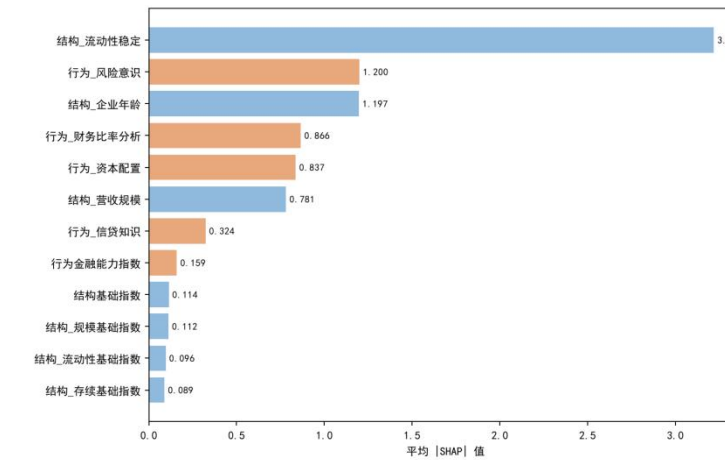


图 4 SHAP 特征重要性

依赖图显示，企业年龄对困境概率呈非单调影响，极年轻与偏成熟区段的风险形态存在差异，单一阈值规则难以充分刻画。个案层面的瀑布图表明，在高风险样本中，流动性指标与风险意识通常同向推高困境概率；在低风险样本中，较强的现金流管理与财务比率分析能力则形成对冲。相较于单一评分，此类个案级归因更适合作为授信决策的解释性附件，也有助于向企业反馈具体的改进方向。

生命周期分层提供了另一维度的证据，如表 4、图 5 所示。表 4 全样本行与表 1 主模型 Ligh

tGBM 使用同一测试集上的同一组预测概率，AUC 为 0.993、KS 为 0.909；分组 KS 刻画的是子样本内部的分层能力，并不等于全样本 KS 的加权平均。各子样本的 AUC 稳定在 0.991—0.995 之间，差异有限；但 KS 呈现出更清晰的模式：初创期为 0.934、成长期为 0.945，均高于成熟期的 0.907。KS 衡量正负样本累积分布的最大间距，反映的是分层能力而非整体排序能力；年轻与成长期企业的行为差异较大、标签相对清晰，模型更易实现分层，而成熟企业的财务模式趋同、困境信号更为隐蔽，分层难度相应上升。这与翁舟杰^[9] 关于市场结构与信贷技术相匹配的讨论并不矛盾：生命周期并非报表分组的形式性划分，而是影响判别力的调节变量。若对不同存续期的企业适用同一套权重，可能在两端同时产生偏误——对年轻企业过度审慎，对成熟企业过度宽松。更为合理的做法，是在统一的模型框架下进行分段校准：对初创与成长期适当提高行为软信息的权重，对成熟期强化流动性与杠杆约束。本文的分层结果为此提供了初步的量化依据，其有效性有待在真实银行样本上进一步检验。

表 4 生命周期分样本模型表现（主模型 LightGBM；全样本与表 1 同一测试集）

分组	测试样本量	AUC	KS
全样本	3,022	0.993	0.909
初创期 (≤5 年)	580	0.995	0.934
成长期 (5—10 年)	581	0.991	0.945
成熟期 (>10 年)	1,861	0.993	0.907

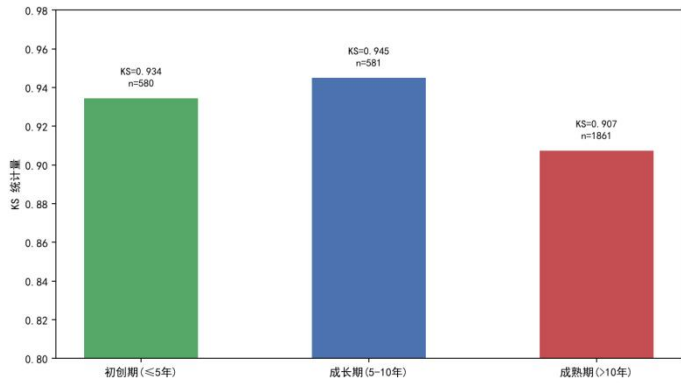


图 5 生命周期异质性

五、讨论与结语

对接近上限的 AUC 需作审慎解读。本文使用的是依据公开 schema 与分布重建的数据，特征与财务困境标签在同一潜变量框架下生成，构念之间的重叠会抬高判别精度的绝对水平；本文

通过训练—测试分离、交叉验证与消融设计尽量抑制乐观偏差，但这些手段不能替代真实违约样本的外部检验。相对结论因而更为可靠：在同一划分规则下，融合维度优于单一维度，行为维度本身即具备可观的判别力，集成方法优于单棵决策树，且 SHAP 归因顺序稳定——这些判断不依赖于绝对精度的高低。因此，本文的价值主要体现在指标构建、对照设计与解释方法层面，而非将 0.993 视为可在任意场景复现的固定水平。信用模型进入实际应用后仍将面临样本漂移、政策冲击与道德风险，公开数据上的高精度只能说明方法思路可行，尚不足以证明其在生产环境中的有效性。

本文在方法上仍存在若干局限。其一，行为变量来源于问卷自评，存在测量误差与社会期许偏差，理想情况下应与银行流水、税务与征信等客观数据交叉验证，本文所用的行为代理属过渡性方案。其二，SHAP 存在固有边界，共线变量会导致归因分摊^[2]；但在现有工具中，它仍是将模型输出转化为审批语言的可行路径——模型给出评分，个案归因进入尽职调查材料，困境概率落于灰区（如 35%—60%）的样本触发人工复核，而非自动拒绝。由此，算法负责排序、人工负责裁量，责任边界更为清晰。本文同步开发了交互式原型系统，用于在非生产环境中演示单户评分与因子分解；其意义不在于直接上线，而在于将指标映射落实为可操作的界面，便于业务人员检验变量设定与解释逻辑的合理性。

参考文献：

- [1] OHLSON J A. Financial ratios and the probabilistic prediction of bankruptcy[J]. Journal of Accounting Research, 1980, 18(1): 109-131.
- [2] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]//Advances in Neural Information Processing Systems, 2017, 30: 4765-4774.
- [3] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.
- [4] KE G, MENG Q, FINLEY T, et al. LightGBM: A highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems, 2017, 30: 3149-3157.
- [5] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32. <https://doi.org/10.1023/A:1010933404324>
- [6] 王鑫, 王莹. 基于 LSTM-CNN 的中小企业信用风险预测[J]. 系统科学与数学, 2022, 42(10): 2698-2711.
- [7] LI J, YAN S, TIAN X, et al. Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data[J]. Transportation Research Part E: Logistics and Transportation Review, 2022, 158: 102611.
- [8] 彭红枫, 叶永刚. 基于还款能力和还款意愿的贷款定价模型研究[J]. 中国管理科学, 2011, 19(6): 40-48.

本文的政策与业务含义可归纳如下。对银行与担保机构而言，中小企业评级不宜局限于规模与抵押，宜在结构基础之外系统采集风险管理制、预算与现金流管理、财务分析工具运用等行为变量并按统一口径入库——消融实验已表明该类信息的边际贡献真实且可验证；SHAP 归因报告可作为授信决策的解释性附件^[12]。对企业自身而言，模型所揭示的改进路径较为明确：应优先夯实流动性与现金流管理，再谋求规模扩张；风险意识与财务比率分析能力在 SHAP 排序中位居前列，可作为可计量的信用因子纳入自我诊断。对监管部门而言，当模型判别精度普遍偏高时，评估重心宜转向校准、分层与可解释性，而非将机器学习方法的采用本身视为创新，以避免出现精度较高但难以落地的情形。

后续研究应在本土银行样本上复现这一框架：以真实逾期标签替代财务困境，以客观流水与征信数据替代行为自评，并沿用同一套消融与 SHAP 流程检验结论的稳健性。该检验的结果将决定本文是一项可迁移的方法研究，还是一次基于公开数据的方法演示。就现有证据而言，本文至少表明：将行为金融能力由控制变量提升为与结构基础并列的维度，在数据上具有可支持性；机器学习相对 Logistic 回归的增益有限但真实；软信息可被量化并加以解释。这为缓解中小企业信贷评估中的信息不对称提供了一条可操作的技术路径。

- [9] 翁舟杰. 关系型贷款、市场结构与小额贷款公司使命漂移[J]. 管理科学学报, 2018, 21(4): 102-113.
- [10] LUSARDI A, MITCHELL O S. The economic importance of financial literacy: Theory and evidence[J]. Journal of Economic Literature, 2014, 52(1): 5-44.
- [11] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [12] BRACKE P, DATTA A, JUNG C, et al. Machine learning explainability in finance: An application to default risk analysis[R]. Bank of England Staff Working Paper No. 816, 2019.