面向不平衡信贷违约预测的改进流模型

卢新月，杨广仁

（暨南大学经济学院，广州 510000）

摘要：针对信贷违约预测中少数类样本分布复杂和类别不平衡问题，提出一种基于 SHAP（沙普利加性解释）引导的混合高斯先验条件 RealNVP（实值非体积保持）模型（SMG-RealNVP）。该模型通过在潜空间中引入混合高斯先验，并结合模态信息构建条件耦合机制，以增强模型对不同违约风险模式的分布建模能力。同时，利用 SHAP 特征重要性设计特征感知掩码，引导耦合层关注关键风险变量，从而提升生成样本的风险语义一致性。并结合多教师知识蒸馏提升轻量化学生模型在不平衡场景下的违约识别能力。在 3 个公开数据集上的实验结果表明，所提方法在 ROC-AUC（受试者工作特征曲线下面积）、AUPRC（精确率-召回率曲线下面积）等关键指标上优于 SMOTE（合成少数类过采样技术）、CTGAN（条件表格生成对抗网络）等基线模型，能够有效捕捉少数类样本的复杂分布，为信贷风险管理提供了一种可落地的技术解决方案。

关键词：信贷违约；类别不平衡；RealNVP（实值非体积保持）；SHAP（沙普利加性解释）；知识蒸馏

中图分类号：F830.5；TP181 **文献标志码**：A **文章编号**：1671－1807(2026)17－0129－10

近年来，随着数字金融、互联网信贷与智能风控的发展，金融机构对信用风险识别的准确性与可解释性提出了更高要求。信用违约预测作为金融风险管理中的核心任务，广泛应用于贷前审批、额度管理与风险定价等场景，对降低违约损失与优化信贷资源配置具有重要意义。然而，真实信贷数据中违约样本占比较低，模型训练过程易偏向多数类正常样本，从而削弱对违约客户的识别能力。同时，不同违约客户在收入结构及负债水平等方面往往存在差异，使少数类样本呈现较强异质性。因此，如何在类别不平衡条件下有效学习违约样本分布特征，并提升模型对高风险客户的识别能力，成为当前智能金融风控研究的重要问题。

针对信贷风险评估，研究主要依赖逻辑回归、判别分析等统计模型[1-2]。Martin[3]将 Logit 回归应用于银行风险预警研究，验证了其在潜在风险识别中的有效性。Laitinen[4]将逻辑回归模型应用于企业信用风险评估。Altman[5]基于财务比率和多元判别分析构建 Z-score 模型，成为企业破产预测的经典方法。随着金融数据复杂度不断提升，传统统计模型在复杂非线性关系建模方面存在一定局限，机器学习方法逐渐被应用于信用违约预测研究。Amzile 和 Habachi[6]将支持向量机（support vector machine，SVM）应用于银行客户违约预测和信用评级，发现径向基核 SVM 较逻辑回归具有更好的预测性能。Wahab 等[7]比较了多种机器学习和深度学习方法在信用卡违约预测中的表现，验证了决策树等模型的应用价值。进一步地，Zhang[8]比较了 XGBoost、Gradient Boosting、LightGBM 等集成模型在贷款违约预测中的表现，表明集成学习能够有效提升复杂信贷数据下的违约预测能力。杨莲和石宝峰[9]结合 ADASYN（自适应合成过采样方法）与 Focal Loss 构建信用风险评价模型，提升了困难违约样本的识别能力。Ying 等[10]构建融合注意力机制的 LightGBM 框架，提升了数字金融场景下信用风险评估效果。与此同时，随着金融监管对模型透明性和可解释性的要求不断提高，可解释人工智能方法逐渐被应用于信用风险评估。Talaat 等[11]将可解释人工智能与深度学习模型结合用于信用卡违约预测，Hlongwane 等[12]基于 Shapley values 构建可解释信用评分框架，增强了信用风险模型的透明性。但现有可解释性研究多侧重于模型后验解释，对关键风险特征在少数类样本生成过程中的引导作用关注不足。

在金融违约预测场景中，类别不平衡问题仍是影响模型识别能力的重要因素。为缓解该问题，现

收稿日期：2026-05-14

作者简介：卢新月（2001—），女，安徽滁州人，硕士研究生，研究方向为机器学习；通信作者杨广仁（1976—），男，山东嘉祥人，博士，教授，研究方向为高维数据的统计推断。

有研究主要从数据层面和算法层面进行改进[13-14]。其中，基于重采样的数据层方法应用最为广泛。Chawla 等[15]提出 SMOTE(合成少数类过采样技术)，通过在少数类样本邻域内进行线性插值生成新样本，是应用最为广泛的过采样方法之一。Han 等[16]在此基础上提出 Borderline-SMOTE，强调对靠近分类边界的少数类样本进行增强。He 等[17]进一步提出 ADASYN，根据少数类样本的学习难度自适应生成合成样本，提高了对难分类区域的关注。这类方法能够在一定程度上缓解类别不平衡问题，但多数仍依赖邻域插值，难以充分刻画高维金融表格数据中的复杂非线性分布及潜在风险子群结构。为突破传统重采样方法的局限，深度生成模型逐渐被引入不平衡数据增强任务。Goodfellow 等[18]提出生成对抗网络(GAN)，为深度生成建模奠定了基础。随后，CTGAN(条件表格生成对抗网络)通过条件采样与模式特定归一化机制，提升了混合类型表格数据的生成能力，成为表格数据生成研究中的代表性方法之一[19]。此外，变分自编码器(VAE)[20]和扩散模型[21]也被应用于数据增强任务。Kotelnikov 等[22]提出的 TabDDPM 将扩散模型扩展至表格数据生成并取得较好生成效果。然而，GAN 类方法易出现模式崩塌与训练不稳定问题，VAE 生成样本易出现过平滑现象，扩散模型则面临计算开销较高的问题。相比之下，归一化流模型(normalizing flow)具有可逆结构、显式概率建模以及训练稳定性等优势，在复杂分布学习任务中展现出较好潜力[23]。其中，NICE(non-linear independent components estimation)[24]、RealNVP(real non-volume preserving)[25]和 Glow[26]等模型能够实现精确似然估计，并在复杂数据分布建模方面取得较好效果。RealNVP 作为典型归一化流模型，具有训练稳定和似然可计算等特点，适合用于少数类样本分布建模。然而，标准 RealNVP 通常采用单高斯先验，难以充分表达金融违约样本中可能存在的多种风险子群结构。已有研究尝试将混合高斯分布引入归一化流模型[27-28]，以增强潜空间对少数类分布的表达能力。

与一般表格数据生成任务不同，金融违约数据往往具有较强的异质性、类别不平衡性以及监管可解释性要求，仅依赖通用生成模型难以同时兼顾生成质量、风险语义一致性与下游风险识别性能。为此，提出一种 SHAP((SHapley additive exPlanations))引导的混合高斯条件 RealNVP 框架，以 RealNVP为基础，引入混合高斯先验以增强潜空间对多风险子群结构的表达能力，并结合条件耦合机制实现类别约束下的稳定生成；同时，利用 SHAP 特征重要性设计特征感知掩码，引导耦合层关注关键风险变量，从而提升生成样本的风险语义一致性。在此基础上，结合多教师知识蒸馏策略，使生成增强与下游分类任务协同优化，以提升模型对少数类违约客户的识别能力。

1 方法框架

1.1 归一化流

归一化流通过一系列可逆变换将复杂数据分布映射到简单的潜空间先验，从而实现显式概率密度建模。相比传统生成模型，其不仅能够精确计算样本似然，还具备稳定的可逆采样能力，因此在复杂分布建模任务中具有较好的适用性。

设原始样本为 x，经可逆双射 f 映射为潜变量 $z=f(x)$。由概率密度的变量变换公式，样本 x 的概率密度为

$$p_x(x)=p_z[f(x)]\cdot\left|\det\left[\frac{\partial f(x)}{\partial x}\right]\right| \tag{1}$$

取对数形式为

$$\ln p_x(x)=\ln p_z[f(x)]+\ln\left|\det\left[\frac{\partial f(x)}{\partial x}\right]\right| \tag{2}$$

式中：$p_x(x)$ 为样本 x 在原始空间中的概率密度；$p_z(z)$ 为潜变量的先验分布；$\det\left[\frac{\partial f(x)}{\partial x}\right]$ 为变换 f 的雅可比矩阵行列式。训练时通过最大化训练样本的对数似然估计模型参数，等价于最小化负对数似然损失：

$$L_{\mathrm{NLL}}=-\frac{1}{N}\sum_{i=1}^{N}\ln p_x(x_i) \tag{3}$$

式中：N 为训练样本数量。

1.2 RealNVP 流模型

在归一化流模型中，RealNVP 通过仿射耦合层构造可逆映射，并简化雅可比行列式求解，从而兼顾复杂分布建模能力与高效似然估计。标准仿射耦合层将输入向量 $\boldsymbol{x}\in\mathbf{R}^D$ 划分为两部分 $\boldsymbol{x}=(\boldsymbol{x}_{1:d},\boldsymbol{x}_{(d+1):D})$，其变换形式为

$$\boldsymbol{y}_{1:d}=\boldsymbol{x}_{1:d} \tag{4}$$

$$\boldsymbol{y}_{(d+1):D}=\boldsymbol{x}_{(d+1):D}\odot\exp[s(\boldsymbol{x}_{1:d})]+t(\boldsymbol{x}_{1:d}) \tag{5}$$

式中：$s(\cdot)$ 和 $t(\cdot)$ 分别为缩放和平移函数；$\odot$ 为逐元素相乘。由于上述结构是可逆的，因此由输出 $\boldsymbol{y}$ 可恢复输入 $\boldsymbol{x}$：

$$\boldsymbol{x}_{1:d}=\boldsymbol{y}_{1:d} \tag{6}$$

$$\boldsymbol{x}_{(d+1):D} = [\boldsymbol{y}_{(d+1):D} - t(\boldsymbol{y}_{1:d})] \odot \exp[-s(\boldsymbol{y}_{1:d})] \tag{7}$$

除可逆性外,RealNVP 的另一个关键优势在于其雅可比行列式为下三角分块结构,可以被高效计算,有

$$\det \boldsymbol{J} = \exp\left[\sum_{j}^{D-d} s_j(\boldsymbol{x}_{1:d})\right] \tag{8}$$

其对数形式为

$$\ln|\det \boldsymbol{J}| = \sum_{j}^{D-d} s_j(\boldsymbol{x}_{1:d}) \tag{9}$$

式中:$s_j(\boldsymbol{x}_{1:d})$ 为缩放向量 $s(\boldsymbol{x}_{1:d})$ 的第 j 个分量。因此,即便 $s(\cdot)$ 和 $t(\cdot)$ 由复杂神经网络实现,RealNVP仍然能高效计算对数行列式。且通过堆叠多个耦合层,模型可逐步增强对复杂数据分布的表达能力。

1.3 混合高斯先验的条件 RealNVP 流模型

尽管标准 RealNVP 能够有效学习复杂分布,但其通常采用单高斯先验,难以刻画金融违约数据中少数类样本的多模态特征,易限制潜空间表达能力,进而影响少数类样本生成质量。为此,提出一种融合混合高斯先验与 SHAP 特征引导机制的条件 RealNVP 模型(SHAP-Mask- Guided Conditional RealNVP with a Gaussian Mixture Prior,SMG-RealNVP),以学习不同风险模式对应的潜空间结构,对于潜变量 $\boldsymbol{z}$,其混合高斯先验分布的概率密度为

$$p_z(\boldsymbol{z}) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{z};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) \tag{10}$$

式中:K 为高斯分量个数;π_k 为第 k 个分量的权重,满足 $\pi_k \geqslant 0$,$\sum_{k=1}^{K}\pi_k = 1$;$\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$ 分别为第 k 个分量的均值和协方差矩阵。

在此基础上,将潜空间中的模态信息融入仿射耦合变换过程。样本 x 映射为潜变量 z 后,其属于第 k 个高斯分量的后验概率为

$$p(k \mid \boldsymbol{z}) = \frac{\pi_k N(\boldsymbol{z};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{h=1}^{K}\pi_h N(\boldsymbol{z};\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_h)} \tag{11}$$

为保证数值稳定性,通过对数空间的最大后验概率准则确定样本最可能所属的模态标记:

$$k^* = \operatorname{argmax}_k[\ln\pi_k + \ln N(\boldsymbol{z};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)] \tag{12}$$

将 k^* 映射为可学习的模态嵌入向量 $\boldsymbol{e}(k^*) \in \boldsymbol{R}^r$($r$ 为嵌入向量维度),并以此改造标准仿射耦合层。设第 l 层的输入为 $\boldsymbol{x} \in \boldsymbol{R}^D$,输出为 $\boldsymbol{y} \in \boldsymbol{R}^D$,掩码向量为 $\boldsymbol{m}^{(l)} \in \{0,1\}^D$,则输入可划分为条件侧和变换侧:

$$\boldsymbol{x}_{\text{cond}} = \boldsymbol{x} \odot \boldsymbol{m}^{(l)},\ \boldsymbol{x}_{\text{trans}} = \boldsymbol{x} \odot (1 - \boldsymbol{m}^{(l)}) \tag{13}$$

随后将模态嵌入 $\boldsymbol{e}(\boldsymbol{k}^*)$ 拼接至缩放与平移函数的输入中,使变换受模态信息约束:

$$s = s_{\text{net}}^{(l)}[\boldsymbol{x}_{\text{cond}},\boldsymbol{e}(\boldsymbol{k}^*)],\ t = t_{\text{net}}^{(l)}[\boldsymbol{x}_{\text{cond}},\boldsymbol{e}(\boldsymbol{k}^*)] \tag{14}$$

此时,第 l 层的前向变换重构为 $\boldsymbol{y} = \boldsymbol{x}_{\text{cond}} + \boldsymbol{x}_{\text{trans}} \odot \exp(\hat{s}) + \hat{t}$。其中,$\hat{s} = (1 - \boldsymbol{m}^{(l)}) \odot s$,$\hat{t} = (1 - \boldsymbol{m}^{(l)}) \odot t$ 为有效缩放与平移项。由于条件变量 k^* 仅作为附加外部输入,未改变原始耦合层的自回归依赖关系,故该变换依然是严格双射且易于求逆的。

在此结构下,整个映射函数由 L 个条件耦合层复合而成,即

$$\boldsymbol{z} = f(\boldsymbol{x};k^*) = f_L \circ f_{L-1} \circ \cdots \circ f_1(\boldsymbol{x};k^*) \tag{15}$$

结合混合高斯先验,模型的负对数似然损失函数为

$$L_{\text{NLL}} = -\frac{1}{N}\sum_{i=1}^{N}\{\ln\pi_{k_i^*} + \ln N[f(x_i;k_i^*);\ \mu_{k_i^*},\Sigma_{k_i^*}] + \sum_{l=1}^{L}\sum_{j=1}^{D}[1 - m_j^{(l)}]s_{i,j}^{(l)}\} \tag{16}$$

该目标可引导潜变量向高概率的主导模态聚集,同时保持流模型前向传播时整体雅可比矩阵的对数行列式仍为三角结构。

同时,为增强模型对关键风险特征的关注能力,进一步结合 SHAP 特征重要性构造特征感知掩码,引导耦合层实现更加合理的特征划分。首先利用 LightGBM 训练辅助分类器,计算每个样本特征的 SHAP 值 ϕ_{ij},并定义第 j 个特征的全局重要性:

$$I_j = \frac{1}{N}\sum_{i=1}^{N}|\phi_{ij}| \tag{17}$$

根据 I_j 对特征降序排序,得到特征重要性序列 $\boldsymbol{\rho} = (\rho_1,\rho_2,\cdots,\rho_D)$。为避免高重要性特征长期处于变换侧,采用高低交错重排策略生成新序列 $\boldsymbol{\rho}' = (\rho_1,\rho_D,\rho_2,\rho_{D-1},\rho_3,\rho_{D-2},\cdots)$。当特征维度 D 为奇数时,最后一个未配对特征顺序追加至序列末尾。基于重排序列按奇偶位置分配掩码,第 1 层耦合层的掩码为

$$m_j^{(1)} = \begin{cases} 1, & \text{如果 } j \in \{\rho_1',\rho_3',\rho_5',\cdots\} \\ 0, & \text{其他} \end{cases} \tag{18}$$

相邻层则采用互补掩码,即 $m_j^{(l+1)} = 1 - m_j^{(l)}$,实现特征条件侧与变换侧的交替更新,提升模型对关键特征的利用效率。

1.4 基于多教师蒸馏的下游分类优化模块

在金融违约识别任务中，若仅依赖原始 0/1 硬标签训练，易使分类器偏向多数类样本，从而影响违约样本的识别能力。知识蒸馏(knowledge distillation，KD)作为一种有效的判别信息传递范式，能够利用教师模型输出的软概率分布，为轻量化模型提供更丰富的类间相关性与置信度信息。

基于此，进一步引入多教师知识蒸馏机制，通过多教师集成输出稳定可靠的软标签，并采用软硬标签联合监督策略，共同指导学生模型训练，以提升违约风险概率的拟合精度。设教师模型数量为 M，第 m 个教师模型对样本 x_i 的违约预测概率为 $p_m(x_i)$，则集成后的平均软标签为

$$\bar{p}_i = \frac{1}{M}\sum_{m=1}^{M} p_m(x_i) \tag{19}$$

进而对软硬标签进行加权融合，构造连续蒸馏目标：

$$\tilde{y}_i = \alpha y_i + (1-\alpha)\,\bar{p}_i,\ \alpha \in [0,1] \tag{20}$$

式中：y_i 为样本 x_i 的真实标签；α 为融合系数。

针对学生模型采用 LightGBM 架构，实验将分类任务转化为风险概率回归问题，并采用均方误差损失对连续监督目标进行拟合优化：

$$L_{\text{stu}}(\omega) = \frac{1}{N}\sum_{i=1}^{N}[g_\omega(x_i) - \tilde{y}_i]^2 \tag{21}$$

式中：$g_\omega(\cdot)$ 为学生模型；ω 为模型参数。该监督方式可约束模型输出在数值上逼近连续监督信号，且有助于学生模型学习更平滑鲁棒的判别函数，从而提升其在类别不平衡与边界模糊场景下的违约识别稳定性。

2 实验设计

2.1 数据集与预处理

为验证 SMG-RealNVP 模型在不平衡信贷违约预测任务中的性能与泛化能力，实验选取 Kaggle 平台的 3 个公开信贷违约数据集。由表 1 可知，各数据集在样本规模、不平衡程度及特征类型等方面存在差异，能较全面地评估模型在不同金融场景下的有效性与鲁棒性。其中，UCI_Credit_Card 数据集来源于中国台湾地区某商业银行真实信用卡客户数据，具有一定实际业务背景。

表 1 数据集概览

数据集	不平衡比	样本量	特征类型
UCI_Credit_Card	1∶3.52	30 000	纯数值型
train_public 个人贷款违约记录数据	1∶4.94	10 000	混合型
Give Me Some Credit	1∶13.96	150 000	纯数值型

为避免数据泄露，实验首先采用分层抽样法将数据集按 8∶2划分为训练集与测试集，并保持正负样本比例与原始分布一致。在训练集内部，进一步采用 5 折分层交叉验证进行模型训练与对比，所有数据预处理和特征变换操作仅在训练折上拟合，再应用于验证折与测试集。数据预处理阶段，剔除无业务意义的字段，对时间、文本等非结构化字段进行特征重构，对信用等级等有序类别特征进行编码。在缺失值处理方面，直接删除缺失率高于 60% 的特征与零方差特征，对剩余数值型与类别型特征分别采用中位数与众数填充，并对未知类别统一编码。考虑到流模型对输入分布较为敏感，进一步对特征进行标准化与分位数变换，将其映射至近似高斯分布以缓解偏态与重尾分布的影响。在生成样本还原至原始特征空间后，对离散及有序变量执行截断与取整操作，以确保取值符合业务逻辑。特征工程方面，结合金融业务语义构建衍生风险特征以增强违约刻画能力。此外，针对数据集维度的差异，采用差异化特征筛选策略，高维数据集基于 LightGBM 计算的 SHAP 值来保留核心特征，低维数据集则保留全部特征，同时利用 SHAP 信息设计耦合层掩码以兼顾信息保留与冗余削减。

2.2 评价指标

由于不平衡违约预测任务中单一指标难以全面反映模型性能，故实验采用多维评价指标，包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-Score、马修斯相关系数(MCC)、ROC-AUC(受试者工作特征曲线下面积)及 AUPRC(精确率-召回率曲线下面积)。此外，引入 Recall@Top K%(K=10)来衡量模型在有限资源下识别高风险客户的能力。由于违约样本占比较低，实验重点关注少数类样本的识别能力，因此将 AUPRC 作为核心指标，该指标对少数类样本敏感，是评估不平衡分类器性能的重要参考。计算公式如下：

$$\begin{cases} \text{Accuary} = \dfrac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\ \text{Recall} = \dfrac{\text{TP}}{\text{TP}+\text{FN}} \\ \text{Precision} = \dfrac{\text{TP}}{\text{TP}+\text{FP}} \end{cases} \tag{22}$$

$$\text{F1} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{23}$$

$$\text{MCC} = \frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{24}$$

$$\text{Recall@Top}K\% = \frac{\text{前 }K\%\text{ 高危样本中包含的真实违约数}}{\text{总违约数}} \tag{25}$$

式中:TP 为正确分类的违约样本;TN 为正确分类的正常样本;FP 为错误分类的正常样本;FN 为错误分类的违约样本。

2.3 实验设置

为避免过度生成样本引入噪声,实验采用均方根采样策略控制生成规模,即将目标不平衡比设为原始比例的平方根,而非强制平衡类别分布。考虑到分类阈值对不平衡任务性能有较大影响,每折交叉验证中,通过步长为 0.01 的全局搜索,在验证集上选取使 F1 最大的分类阈值,以提高模型评估结果的稳定性。为保证实验可复现性与对比公平性,所有实验均固定随机种子为 42,并采用统一的数据划分与交叉验证策略。

在统一实验设置的基础上,进一步配置模型训练参数。以 Give Me Some Credit 数据集为例,SMG-RealNVP 流模型采用 6 层仿射耦合层,层间插入 ActNorm 层稳定训练;潜变量模态信息通过维度为 16 的可学习嵌入向量,注入耦合层的缩放与平移网络。训练使用 Adam 优化器,初始学习率为 0.005,权重衰减系数为 10^{-5},结合余弦退火调度与梯度裁剪。混合高斯先验分量数 K 通过潜空间 BIC(Bayesian information criterion)自适应确定。训练分为两个阶段,预热阶段仅采用少数类样本学习基础潜空间分布,正式训练过程引入早停机制以防止过拟合。在样本生成阶段,对每个分量采用均匀配额策略生成候选样本,并通过流模型似然和 MLP(multi-layer perceptron)探针剔除低密度和低置信度样本,以保证生成样本与少数类真实分布一致。下游分类模块通过 3 个 LightGBM 教师模型对训练样本进行集成预测以生成软标签,并按权重 $\alpha=0.7$ 将其与真实标签融合以构建连续训练目标。随后学生模型使用 LightGBM 回归器拟合该目标,并在验证集上进行概率校准与阈值选择。为避免生成样本引发特征漂移,将生成样本的权重 w_{gen} 设置为 0.5。

3 实验结果与分析

3.1 实验结果

为评估 SMG-RealNVP 在不平衡违约预测任务中的分类性能与生成质量,实验选取传统过采样方法(SMOTE、ADASYN、Borderline-SMOTE)、深度生成方法(CTGAN)及基准模型作为对比,所有方法均以 LightGBM 为下游分类器,并采用统一的数据划分与交叉验证策略。实验过程中,交叉验证结果整体波动较小,说明模型训练过程较为稳定。表 2 给出了各方法在测试集上的最终评估结果。由表 2 可知,SMG-RealNVP 在 3 个数据集上整体表现较优,在 Give Me Some Credit 与 UCI_Credit _Card 数据集上,AUPRC 与 F1 均略优于其他方法。在 train_ public 数据集上,SMG-RealNVP 的 ROC-AUC、AUPRC、F1 及 MCC 均保持相对较高水平。此外,在业务相关指标 Recall@Top10%上,SMG-RealNVP 在 Give Me Some Credit 上表现较佳,表明其在高风险样本排序方面具有一定优势。

表 2 测试集结果

数据集	方法	ROC-AUC	AUPRC	F1	Accuracy	Recall	MCC	Recall@Top10%
Give Me Some Credit	Baseline	0.860 2	0.330 7	0.399 1	0.873 4	0.628 9	0.370 0	0.523 2
	SMOTE	0.865 1	0.380 4	0.420 9	0.888 3	0.607 5	0.388 1	0.554 1
	ADASYN	0.865 0	0.380 2	0.422 8	0.889 3	0.606 5	0.389 8	0.547 6
	Borderline-SMOTE	0.865 3	0.375 0	0.416 1	0.882 7	0.625 4	0.385 9	0.541 6
	CTGAN	0.869 2	0.390 3	0.426 3	0.889 4	0.615 0	0.394 4	0.553 6
	SMG-RealNVP	0.869 4	0.402 4	0.430 4	0.891 2	0.615 0	0.398 3	0.557 1
train_public	Baseline	0.870 8	0.515 6	0.573 6	0.801 5	0.792 3	0.487 9	0.350 1
	SMOTE	0.871 1	0.528 5	0.573 3	0.809 5	0.759 6	0.484 0	0.335 3
	ADASYN	0.871 6	0.523 0	0.572 7	0.810 5	0.753 7	0.482 6	0.350 1
	Borderline-SMOTE	0.868 3	0.515 9	0.568 7	0.799 0	0.786 4	0.481 4	0.323 4
	CTGAN	0.879 0	0.538 6	0.575 6	0.809 0	0.768 5	0.487 5	0.365 0
	SMG-RealNVP	0.881 1	0.556 7	0.582 2	0.817 0	0.756 7	0.494 3	0.350 1
UCI_Credit_Card	Baseline	0.774 2	0.523 1	0.541 0	0.801 2	0.529 8	0.414 3	0.313 5
	SMOTE	0.775 9	0.55 56	0.528 1	0.784 3	0.545 6	0.388 8	0.313 5
	ADASYN	0.773 2	0.552 3	0.523 7	0.780 5	0.545 6	0.381 9	0.315 0
	Borderline-SMOTE	0.766 6	0.548 4	0.524 6	0.787 3	0.530 5	0.387 7	0.310 5
	CTGAN	0.776 5	0.547 5	0.544 0	0.779 8	0.593 8	0.402 7	0.315 0
	SMG-RealNVP	0.780 7	0.558 7	0.544 0	0.784 3	0.581 8	0.405 0	0.312 7

为分析模型生成样本的分布保真性，实验选取关键业务特征绘制真实少数类样本与生成样本的核密度估计曲线（图 1）。由图 1 可知，对于年龄近似单峰特征，生成样本较好地保留了主峰位置和整体变化趋势；对于人均收入这类多峰特征，模型能够捕捉主要分布结构，但局部峰高度略有偏差；对于具有尖峰与长尾特征的循环信用利用率，生成样本在主体区间和尾部区间均与真实样本保持较高一致性。总体来看，生成样本在保留原始特征和边际统计特征的同时实现适度分布扩增，而非简单复制真实样本，为下游分类器提供了丰富且高保真的训练数据。

此外，为验证模型对关键风险特征的识别效果及其可解释性，进一步对下游分类模型进行 SHAP 归因分析。由图 2 可知，循环信用利用率和逾期 90 d 以上次数是影响模型输出的重要特征。其中，逾期次数增加通常对应更高的正向 SHAP 值，而年龄较低的样本也表现出更高的违约风险，与实际业务认知基本一致。这说明 SHAP 特征感知掩码能够有效强化模型对关键风险特征的建模能力，从而提升生成样本的风险语义一致性与模型可解释性。

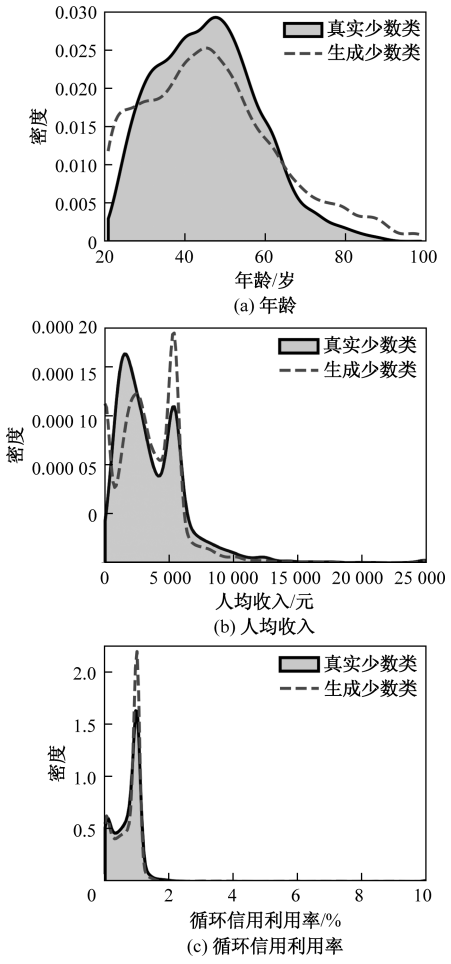


图 1　真实与生成少数类样本的核密度估计曲线

3.2　消融实验

为验证各核心模块对模型性能的贡献，实验在 Give Me Some Credit 数据集上开展消融实验。以标准 RealNVP 为基础，分别引入混合高斯先验与条件耦合机制（G）、SHAP 特征引导机制（S）以及知识蒸馏模块（KD），并进一步构建不同模块组合以分析各模块的独立作用及协同增强效果。由表 3 可知，仅采用标准 RealNVP 时（M1），模型在少数类违约样本识别方面仍存在一定局限，AUPRC 和 F1 分别仅为 0.351 2 和 0.397 5。引入混合高斯先验与条件耦合机制后（M2），ROC-AUC、AUPRC 及 F1 均有所提升，说明该机制能够增强模型对少数类复杂多模态分布的表达能力。仅加入 SHAP 特征引导机制时（M3），模型在 F1 与 MCC 指标上较基线模型进一步提升，表明通过关注关键风险变量，模型对少数类违约特征的表达能力得到增强。相比之下，知识蒸馏模块对模型性能提升最为明显。当仅引入 KD 模块时（M4），ROC-AUC、AUPRC 和 F1 分别提升至 0.870 0、0.406 6 和 0.431 0，Recall@Top10%达到 0.556 6，说明知识蒸馏能够有效优化学生模型决策边界，增强对高风险违约客户的识别能力。在双模块组合实验中，各组合模型整体性能较单模块方案进一步提升，说明不同模块之间具有一定协同增强作用。其中，M6 与 M7 在 AUPRC、F1 及 MCC 等指标上表现较优，表明生成增强、特征引导与蒸馏学习之间能够形成有效互补。综合来看，融合混合高斯先验、SHAP 特征引导及知识蒸馏机制的 SMG-RealNVP 模型（M8）在 ROC-AUC（0.871 3）、AUPRC（0.406 9）及 Recall@Top10%（0.558 1）等指标上取得最优表现，同时在 Accuracy、Recall 及 MCC 等指标上也保持较优水平，说明各模块联合后能够在整体风险区分能力与少数类违约样本识别能力之间取得较好平衡。

3.3　超参数敏感性分析

为确定混合高斯先验的分量数 K，实验在流模型完成预热训练后，将少数类样本映射至潜空间，并利用 BIC 准则对不同分量数的高斯混合模型进行选择。以 Give Me Some Credit 数据集为例，由图 3 可知，BIC 随分量数增加整体下降，但当 $K \geqslant 4$ 时下降幅度明显趋缓，继续增加分量数带来的收益有限。因此，后续实验统一设置 $K = 4$。

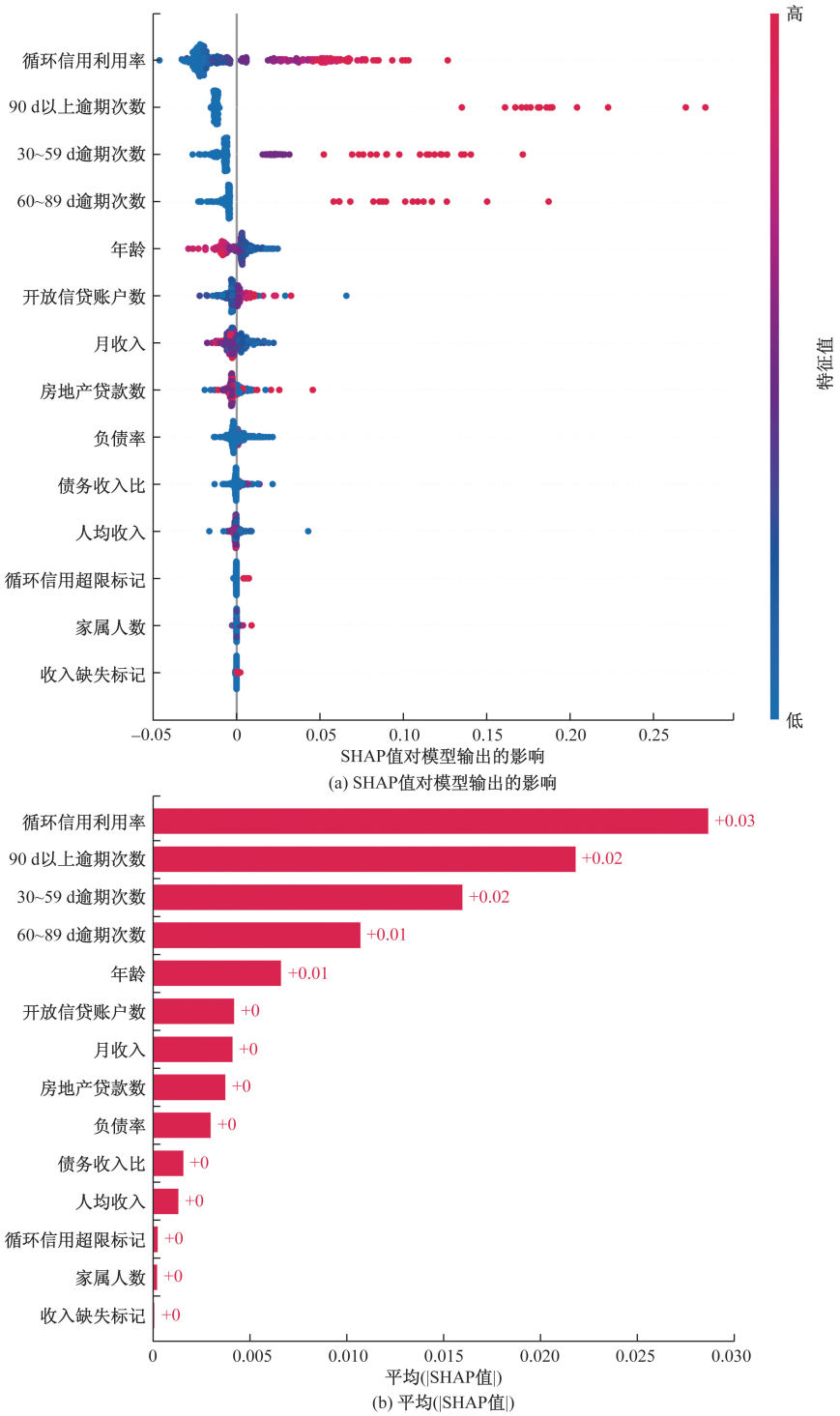


图 2　基于 SHAP 值的模型全局可解释性分析

表 3　消融实验结果

模型	ROC-AUC	AUPRC	F1	Accuracy	Recall	MCC	Recall@Top10%
M1:RealNVP	0.861 0	0.351 2	0.397 5	0.876 3	0.610 5	0.365 8	0.514 2
M2:RealNVP+G	0.863 7	0.377 9	0.408 6	0.881 4	0.613 0	0.376 9	0.536 7
M3:RealNVP+S	0.862 7	0.355 2	0.404 2	0.880 2	0.608 0	0.371 9	0.526 7
M4:RealNVP+KD	0.870 0	0.406 6	0.431 0	0.892 5	0.609 5	0.398 2	0.556 6
M5:RealNVP+G+S	0.866 0	0.367 4	0.415 1	0.884 7	0.612 0	0.383 1	0.539 7
M6:RealNVP+G+KD	0.870 1	0.406 0	0.433 2	0.892 4	0.615 5	0.401 1	0.557 1
M7:RealNVP+S+KD	0.870 4	0.404 9	0.435 1	0.893 8	0.612 0	0.402 5	0.556 1
M8:RealNVP+G+S+KD	0.871 3	0.406 9	0.430 3	0.891 4	0.614 0	0.398 1	0.558 1

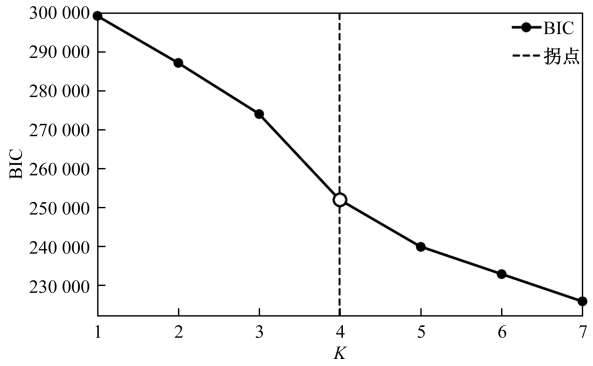


图 3　基于潜空间 BIC 的分量数选择

在固定 $K=4$ 的条件下，进一步对蒸馏权重 α 、生成样本权重 w_{gen} 以及模态嵌入维度 embed_dim 进行敏感性分析。由图 4 可知，模型的 ROC-AUC、AUPRC和 F1 指标随各超参数变化整体波动较小，说明该模型对关键超参数具有较好的稳定性。

4　结论

针对金融违约预测中少数类样本稀缺及分布异质性问题，提出并验证了基于流模型的少数类生成增强方法。通过实验分析与模型验证，得到如下结论。

(1)金融违约少数类样本不仅存在数量不足问题，同时还具有明显的异质性与多模态结构。传统基于局部插值的过采样方法难以充分学习复杂风险模式，而实验所构建的混合高斯条件 RealNVP 框架能够较好地学习少数类样本的复杂分布特征，在保持生成样本质量的同时，有效提升违约样本的识别效果与模型泛化能力，这说明基于显式概率建模的流模型在金融不平衡表格数据场景下具有较好的应用潜力。

(2)通过 SHAP 特征重要性引导生成过程，模型在关注整体数据分布的同时，能够更加关注关键风险特征，从而增强生成样本的风险语义一致性与结果可解释性。结果表明，将特征重要性信息与生成建模过程相结合，有助于提升模型在金融风控场景下的业务适配能力，也为可解释智能风控模型的构建提供了新的思路。

(3)从实际应用角度看，该方法能够在一定程度上缓解金融机构在贷前审批与风险管理过程中因违约样本稀缺导致的识别偏差问题，对于提升高风险客户识别能力与风险预警水平具有一定参考价值。在当前金融风控日益强调模型透明性、风险识别可靠性与审慎监管要求的背景下，兼顾生成质量与业务可解释性的少数类增强方法，有助于提升

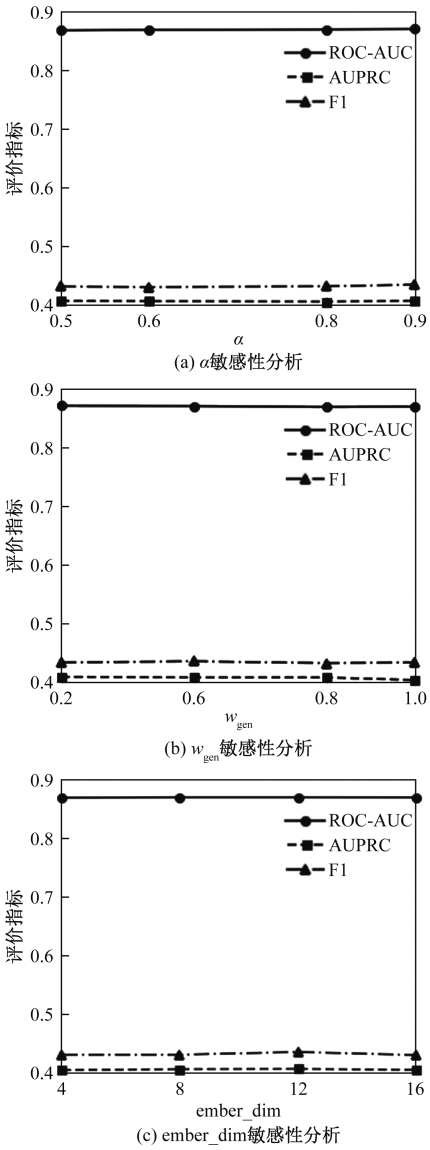


图 4　超参数敏感性分析

智能风控模型在复杂金融场景中的稳定性与可信度，也可为金融机构开展精细化风险管理与智能化风险决策提供一定参考。

(4)尽管该方法在金融违约预测任务中取得了较好的实验效果，但在复杂金融表格数据场景下仍有进一步研究空间。例如，流模型在混合型表格数据建模方面仍有改进空间，尤其是在连续、序数与离散特征的统一表示方面。且未来还可研究在噪声约束下合理引入边界样本，以增强模型对边缘风险模式的刻画能力。

参考文献

[1] HAND D J, HENLEY W E. Statistical classification methods in consumer credit scoring: a review[J]. Journal of the Royal Statistical Society: SeriesA (Statistics in So-

ciety), 1997, 160(3): 523-541.

[2] THOMAS L C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers [J]. International Journal of Forecasting, 2000, 16(2): 149-172.

[3] MARTIN D. Early warning of bank failure: alogit regression approach[J]. Journal of Banking & Finance, 1977, 1(3): 249-276.

[4] LAITINEN E K. Predicting a corporate credit analyst's risk estimate by logistic and linear models[J]. International Review of Financial Analysis, 1999, 8(2): 97-121.

[5] ALTMAN E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. The Journal of Finance, 1968, 23(4): 589-609.

[6] AMZILE K, HABACHI M. Assessment of Support Vector Machine performance for default prediction and credit-rating[J]. Banks and Bank Systems, 2022, 17(1): 161-175.

[7] WAHAB F, JIANG H, SANI N S, et al. Credit card default prediction using ML and DL techniques[J]. Intelligent Systems with Applications, 2024, 4: 293-306.

[8] ZHANG X. Data-driven loan default prediction: a machine learning approach for enhancing business process management[J]. Systems, 2025, 13(7): 581.

[9] 杨莲, 石宝峰. 基于Focal Loss修正交叉熵损失函数的信用风险评价模型及实证[J]. 中国管理科学, 2022, 30(5): 65-75.

[10] YING C W, SHI A L, LI X Y. Hybrid boosted attention-based LightGBM framework for enhanced credit risk assessment in digital finance[J]. Humanities and Social Sciences Communications, 2025, 12(1): 1-13.

[11] TALAAT F M, ALJADANI A, BADAWY M, et al. Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction[J]. Neural Computing and Applications, 2024, 36(9): 4847-4865.

[12] HLONGWANE R, RAMABOA K, MONGWE W. A novel framework for enhancing transparency in credit scoring: leveraging Shapley values for interpretable credit scorecards[J]. PLoS ONE, 2024, 19(8): e0308718.

[13] HE H, GARCIA E A. Learning from imbalanced data [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.

[14] GUO H, LI Y, SHANG J, et al. Learning from class-imbalanced data: review of methods and applications[J]. Expert Systems with Applications, 2017, 73: 220-239.

[15] CHAWLA N, BOWYER K, HALL L O, et al. SMOTE: synthetic minority over-samplingtechnique[J]. Journal of Artificial Intelligence Research, 2002, 16(1): 321-357.

[16] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets[C]//Lecture Notes in Computer Science: Advances in Intelligent Computing-ICIC 2005. Berlin: Springer, 2005, 3646: 878-887.

[17] HE H, BAI Y, GARCIA E A, et al. ADASYN: adaptive synthetic sampling approach for imbalanced learning [C]//2008 IEEE International Joint Conference on Neural Networks (IJCNN). Hong Kong: IEEE, 2008:1322-1328.

[18] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2014, 63: 139-144.

[19] XU L, SKOULARIDOU M, CUESTA-INFANTE A, et al. Modeling tabular data using conditional GAN [C]//Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Vancouver: Curran Associates, 2019: 7335-7345.

[20] KINGMA D P, WELLING M. Auto-encodingvariational Bayes[EB/OL]. (2013-12-20)[2026-04-30]. https://doi.org/10.48550/arXiv.1312.6114.

[21] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[22] KOTELNIKOV A, BARANCHUK D, RUBACHEV I, et al. TabDDPM: modelling tabular data with diffusion models[C]//Proceedings of the 40th International Conference on Machine Learning (ICML 2023). Honolulu: PMLR, 2023: 17564-17579.

[23] PAPAMAKARIOS G, NALISNICK E, REZENDE D J, et al. Normalizing flows for probabilistic modeling and inference[J]. Journal of Machine Learning Research, 2021, 22(57): 1-64.

[24] DINH L, KRUEGER D, BENGIO Y. NICE: non-linear independent components estimation[C]//Proceedings of the 3rd International Conference on Learning Representations Workshop (ICLR). San Diego: ICLR, 2015 (arXiv:1410.8516).

[25] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using Real NVP[C]//Proceedings of the 5th International Conference on Learning Representations (ICLR). Toulon: ICLR, 2017 (arXiv:1605.08803).

[26] KINGMA D P, DHARIWAL P. Glow: generative flow with invertible 1×1 convolutions[C]//Advances in Neural Information Processing Systems 31 (NeurIPS 2018). Montreal: Curran Associates, 2018: 10215-10224.

[27] IZMAILOV P, KIRICHENKO P, FINZI M, et al. Semi-supervised learning with normalizing flows[C]//Proceedings of the 37th International Conference on Machine Learning (ICML 2020). PMLR, 2020, 119: 4615-4630.

[28] HO J, CHEN X, SRINIVAS A, et al. Flow++: improving flow-based generative models with factored base distributions and continuous updates[C]//Proceedings of the 36th International Conference on Machine Learning (ICML 2019). PMLR, 2019, 97: 2722-2730.

An Improved Flow-based Model for Imbalanced Credit Default Prediction

LU Xinyue, YANG Guangren

(School of Economics, Jinan University, Guangzhou 510000, China)

Abstract: In credit default prediction, complex distributions are often observed in minority-class samples, leading to severe class imbalance. To address these issues, a SHAP-Mask-Guided Conditional RealNVP model with a Gaussian Mixture Prior (SMG-RealNVP) was proposed. A Gaussian mixture prior was introduced into the latent space, and modality information was incorporated to construct conditional coupling mechanisms. Consequently, the model's capability to capture distributional patterns of heterogeneous default risk modes was enhanced. In addition, feature-aware masks were constructed based on SHAP feature importance to guide the coupling layers toward focusing on critical risk-related variables. As a result, the risk semantic consistency of the generated samples was improved. Moreover, a multi-teacher knowledge distillation strategy was employed to enhance the default identification ability of a lightweight student model under class imbalance. From the experimental results on three public datasets, it is found that the SMG-RealNVP model is superior to baseline methods such as SMOTE and CTGAN on key metrics including ROC-AUC and AUPRC. It is also found that the complex distributions of minority-class samples are effectively captured, and a practical technical solution for credit risk management is provided.

Keywords: credit default; class imbalance; RealNVP (Real-valued Non-Volume Preserving); SHAP (SHapley Additive exPlanations); knowledge distillation