Management Science and Engineering 管理科学与工程, 2025, 14(1), 67-77
Published Online January 2025 in Hans. https://www.hanspub.org/journal/mse
https://doi.org/10.12677/mse.2025.141008

Hans汉斯

基于高斯加权局部异常因子过滤的成本敏感信用评分模型研究

王全东*，郭　楷，范　宏#

东华大学旭日工商管理学院，上海

收稿日期：2024年12月9日；录用日期：2025年1月4日；发布日期：2025年1月13日

摘　要

信用评分作为金融风险管理和决策制定的核心环节，对于金融机构的稳健运营和市场竞争力至关重要。然而，信贷数据中普遍存在的类别不平衡现象，即违约样本数量远小于非违约样本数量，给信用评分模型的构建带来了挑战，容易导致模型偏向多数类而忽略少数类，从而降低模型的预测准确性和泛化能力。为解决这一问题，提出的GLOF-BFL-LightGBM模型采用了一种分阶段的优化策略。首先，考虑到异常样本的存在会进一步加剧类别不平衡的影响，并降低模型的鲁棒性，本研究引入高斯加权局部异常因子(GLOF)技术，识别并剔除数据中的潜在异常样本，以净化数据集并提高模型的稳定性。其次，为了提升模型对少数类的识别能力，采用Focal Loss损失函数来降低多数类样本对模型训练的影响，并利用贝叶斯优化技术自动搜索Focal Loss损失函数的最优参数，以获得最佳的类别不平衡学习效果。为验证模型的有效性，本文在UCI数据库的四个信贷数据集上进行了实验，并将GLOF-BFL-LightGBM模型与多种基线模型(包括传统的分类方法和常规的集成学习模型)进行了比较。实验结果表明，GLOF-BFL-LightGBM模型在AUC、KS值等关键指标上均优于对比模型，有效提升了信用评分的准确性和模型的泛化能力，为个人信用风险评估提供了一种可靠的工具。

关键词

信用评分，类不平衡，集成学习，LightGBM

Research on Cost-Sensitive Credit Scoring Model Based on Gaussian-Weighted Local Anomaly Factor Filtering

Quandong Wang*, Kai Guo, Hong Fan#

Glorious Sun School of Business and Management, Donghua University, Shanghai

*第一作者。
#通讯作者。

文章引用: 王全东, 郭楷, 范宏. 基于高斯加权局部异常因子过滤的成本敏感信用评分模型研究[J]. 管理科学与工程, 2025, 14(1): 67-77. DOI: 10.12677/mse.2025.141008

Received: Dec. 9th, 2024; accepted: Jan. 4th, 2025; published: Jan. 13th, 2025

Abstract

Credit scoring, as a core aspect of financial risk management and decision making, is crucial to the sound operation and market competitiveness of financial institutions. However, the prevalent category imbalance in credit data, in which the number of default samples is much smaller than the number of non-default samples, poses a challenge to the construction of credit scoring models, which can easily lead to a model biased toward the majority class and ignoring the minority class, thus reducing the predictive accuracy and generalization ability of the model. To address this problem, the proposed GLOF-BFL-LightGBM model adopts a staged optimization strategy. First, considering that the presence of anomalous samples can further exacerbate the effect of category imbalance and reduce the robustness of the model, this study introduces the Gaussian-weighted local anomaly factor (GLOF) technique to identify and remove potentially anomalous samples from the data in order to purify the dataset and improve the stability of the model. Secondly, in order to improve the model's ability to identify the minority class, the Focal Loss loss function is used to reduce the impact of the majority class samples on the model training, and Bayesian optimization technique is used to automatically search for the optimal parameters of the Focal Loss loss function in order to obtain the best class imbalance learning effect. To verify the effectiveness of the model, experiments are conducted on four credit datasets of the UCI database, and the GLOF-BFL-LightGBM model is compared with a variety of baseline models (including traditional classification methods and conventional integrated learning models). The experimental results show that the GLOF-BFL-LightGBM model outperforms the comparison models in key metrics such as AUC and KS values, effectively improves the accuracy of credit scoring and the generalization ability of the model, and provides a reliable tool for personal credit risk assessment.

Keywords

Credit Scoring, Class Imbalance, Integrated Learning, LightGBM

1. 引言

随着现代经济的发展和消费观念的改变，贷款已经成为企业和个人解决资金难题的主要手段。银行和其他金融贷款机构通过信用评分来衡量贷款申请人的信用状况，从而拒绝接受高风险信用贷款的申请者的信贷申请，以减少不良贷款和贷款风险[1]。鉴于信用评分的准确性与所采用的信用评分模型紧密相关，为了增强盈利潜力，对这些模型进行优化显得尤为关键。

信用评分是指对客户的财务状况、信用历史等信息进行综合分析，以构建分类模型，将客户划分为违约和非违约两大类[2]。最原始的信用评分方法是专家主观分析法，即金融机构的专业分析人员依靠5C 原则对申请人是否违约进行主观判断。虽然专家分析法能评估借款人信用，但依赖于风控人员的高业务水平和主观判断，成本较高且不适合大规模数据的分析。随着研究的深入，数学和统计方法逐渐应用于信用评分模型中，解决了专家主观分析法的主观判断影响和高成本问题。Durand 等[3]首先使用判别分析法区分“好”的贷款和“坏”的贷款，定量评估贷款者的信用风险。Orgler 等[4]将线性回归方

法应用于消费信贷的信用评分，将借款人分类到不同的风险等级中。随着计算机技术的发展，机器学习算法在信用评分领域得到了迅速发展，进一步提升了模型的预测精度。Zhang 等[5]提出了一种基于最优群智能算法训练的神经网络信用评分模型框架，通过优化超参数，提高了反向传播人工神经网络(BP-ANN)的拟合和泛化能力，从而提升了模型的预测精度。Zhou 等[6]一种基于模糊积分的集成支持向量机来解决信用风险评估问题，提高预测精度的同时减少计算成本和内存占用。单个机器学习模型的分类性能有限，集成学习逐渐应用到信用评分领域。王重仁等[7]提出了一种基于贝叶斯参数优化和XGBoost 算法的信用评估方法，实验证实预测效果优于其他树集成模型算法。Liu 等[8]提出了一种异构集成和加权投票机制的 Heter-DF 模型，有效提升大规模和小规模信用评分的准确性与鲁棒性。康海燕等[9]利用 Stacking 集成学习技术构建个人信用评分模型，在某电信数据验证出模型具有很好的预测能力。

通常情况下，信贷建模数据往往呈现不平衡的类别分布，即非违约客户样本数量较多，而违约客户样本数量较少。传统的分类模型往往会偏向于多数类样本(非违约客户)的特征，导致少数类样本(违约客户)难以被正确识别。如果继续采用传统模型对客户进行分类，整体分类性能可能会下降。信用评分领域中正负样本比例失衡问题是制约模型性能的关键因素之一。现有研究主要从算法层面和数据层面两种策略入手缓解该问题。算法调整方法主要指代价敏感学习，通过在模型训练过程中引入错误分类的代价概念，使得模型能够根据各类错误带来的不同后果进行优化，以最小化总体代价。Nasser 等[10]提出了一种新的混合性能度量方法，用于实现成本敏感的信用评分，实验结果显示该模型具备良好的分类性能。Wu 等[11]提出了一个面向不确定性的成本敏感信用评分框架，并进行了多目标特征选择，定义了基于正确分类到两个类别(好和坏)的概率的成本敏感目标函数，以优化算法的参数。Bahnsen 等[12]提出了一种新的基于实例的成本矩阵，并将实例依赖的成本引入逻辑回归模型中，用于信用评分。数据分布调整方法主要包括过采样、欠采样。过采样方法通过增加少数类的样本数量来平衡类别分布，减少模型对多数类的偏见，从而提高对少数类样本的识别能力。Shen 等[13]提出了一个基于改进的 SMOTE 技术的深度学习集成信用风险评估模型，提高信用评分的准确性。欠采样方法通过减少多数类的样本数量，以降低其在训练数据中的过度表示，从而减少模型对多数类的偏好并提高对少数类的识别。邵良杉等[14]提出了一种基于 Fisher-SDSMOTE-ESBoostSVM 的类别不平衡信用评分预测模型，使用基于支持度的过采样算法(SDSMOTE)处理信贷样本不平衡数据。陈启伟等[15]提出一种基于 Ext-GBDT 集成的类别不平衡信用评分模型，使用欠采样的方法从“好”客户(大类)中随机采样多份与全部“坏”客户(小类)等量的样本，并用不同的训练子集及特征采样和参数扰动的方法训练得到多个差异化的 Ext-GBDT 子模型，通过实验证实了模型的有效性。

本文的贡献主要体现在：(1) 将贝叶斯优化算法与 Focal Loss 损失函数相结合，并嵌入到 LightGBM 树集成模型中。通过贝叶斯优化，可以针对不同数据集动态调整 Focal Loss 函数中的权重系数和聚焦参数，从而自适应地学习类别不平衡数据，并更加关注难分类样本相比于传统的信用评分模型，该模型在识别高风险用户方面表现出更高的准确性和有效性。(2) 提出了一种基于高斯加权局部异常因子(GLOF)的信贷数据预处理方法。该方法通过引入高斯核函数，对局部可达密度进行加权计算，从而更精确地识别并剔除数据集中的极端离群点。该预处理步骤有效降低了 LightGBM 模型过拟合的风险，并避免了 Focal Loss 函数过度关注难以正确分类的极端离群点，从而提升了模型的整体信用评分性能和鲁棒性。(3)不仅拓展了信用评分的理论框架，也为金融风险管理提供了有效的实践工具。所提出的模型通过更精确地识别高风险客户，能够有效支持金融机构在风险定价、授信额度管理等方面的决策优化，进而提升整体风险管理的效率和精准性。此外，模型的构建思路和方法也为其他不平衡数据分类问题的研究提供了借鉴和参考，具有重要的理论意义和应用价值。

2. 理论基础及模型

2.1. LightGBM

LightGBM 是一种基于梯度提升(Gradient Boosting)算法的集成学习方法，它通过构建多个弱学习器(通常是决策树)，并将这些学习器的结果组合成一个更强的预测模型。该方法在模型训练过程中利用梯度信息进行迭代优化，逐步提高模型的预测精度。

LightGBM 使用直方图算法来优化训练过程，它将连续特征值离散化到固定数量的桶(bins)中，每个桶代表一个特征的一个区间。在训练过程中，模型不需要计算每个可能的分裂点，只需要计算每个桶的分裂，从而加快了训练速度。LightGBM 采用叶子优先策略，即每次选择使损失函数最小的叶子节点进行分裂。这种策略可以使树的深度较大，从而提高模型的拟合能力，并且通常可以更有效地降低损失。然而，这种策略的风险是可能会过拟合。

LightGBM 的目标是最小化一个损失函数和正则化项的组合。具体的目标函数通常为：

$$L(\theta)=\sum_{i}^{N} l(y_i,\hat{y}_i)+\sum_{k}^{K}\Omega(f_k) \tag{1}$$

其中 $l(y_i+\hat{y}_i)$ 是损失函数，度量模型预测值和真实标签之间的差距； $\Omega(f_k)$ 是正则化项，控制模型的复杂度； y_i 是第 i 个样本的真实标签； $\hat{y}_i$ 是第 i 个样本的预测输出。

2.2. 贝叶斯优化 FocalLoss

FocalLoss 是一种为了解决类别不平衡问题而设计的损失函数，特别是在目标检测任务中，正负样本数量差异很大时非常有用。FocalLoss 的核心思想是通过引入一个调节因子，减小对易分类样本(如背景类样本)的损失贡献，增加对难分类样本(如少数类样本)的关注，从而提升模型的性能。

对于二分类问题，FocalLoss 的公式如下：

$$FL(p_t)=-\alpha(1-p_t)^{\gamma}\,log(p_t)-(1-\alpha)\,p_t^{\gamma}\,log(1-p_t) \tag{2}$$

其中， p_t 是模型对于每个类别的预测概率； α 是平衡因子，用来控制正负样本的权重； γ 是调制因子，用来调整对易分类样本的关注度，增强对难分类样本的关注。

在实际训练过程中，由于 FocalLoss 引入了 α 平衡因子和 γ 调制因子两个超参数，这些参数的取值对模型性能有较大影响。对于不同的训练数据， α 平衡因子和 γ 调制因子往往需要根据实际样本分布进行针对性调优，降低模型的泛化能力。贝叶斯优化是一种基于贝叶斯定理的全局优化算法，它通过建立一个目标函数的概率模型来指导搜索过程，从而找到使目标函数取得最优值的参数配置。这种方法在较少的迭代次数内就能找到接近最优解的参数配置，并且能够处理多峰、非凸等复杂的目标函数。

对于 α 和 γ 调优，贝叶斯优化的目标是通过最小化验证集上的损失 $\mathcal{L}_{val}(\alpha,\gamma)$，其目标是：

$$\theta^{*}=\arg\min_{\theta=(\alpha,\gamma)}\mathcal{L}_{val}(\alpha,\gamma) \tag{3}$$

其中， $\theta=(\alpha,\gamma)$ 是要优化的超参数集合。

贝叶斯优化中的代理模型通常使用高斯过程(GP)来拟合目标函数。高斯过程的建模公式如下：

$$f(\theta)\sim GP\left(m(\theta),k(\theta,\theta')\right) \tag{4}$$

其中， $f(\theta)$ 是目标函数，表示给定超参数 $\theta=(\alpha,\gamma)$ 时的损失； $m(\theta)$ 是均值函数，通常设为零； $k(\theta,\theta')$ 是协方差函数(核函数)，用于描述超参数之间的相关性。

贝叶斯优化通过采集函数来选择下一个实验的超参数组合。常用的采集函数之一是期望改进

(Expected Improvement, EI)。期望改进的公式为：

$$EI(\theta)=E\left[\max\left(f\left(\theta^{+}\right)-f(\theta),0\right)\right] \tag{5}$$

其中：$f\left(\theta^{+}\right)$是当前已知的最优目标函数值；$f(\theta)$是高斯过程模型在超参数θ下的预测。

贝叶斯优化选择使得期望改进最大的超参数θ，即选择一组新的α和γ，进行下一轮的训练和评估。

2.3. 高斯加权的 LOF 技术

Focal Loss 的设计初衷是通过降低对易分类样本的关注，并增加对难分类样本的关注。虽然这在类别不平衡的情况下有助于提高模型的性能，但如果数据集中存在离群点(outliers)或噪声，模型可能会过度关注这些错误分类的样本。由于离群点并不代表数据的整体分布特征，模型可能会偏离全局的最优解，甚至出现过拟合。因此，对信贷样本数据的离群点的筛除对模型的分类效果至关重要。

本文提出了一种高斯加权调整局部可达密度的 LOF (Local Outlier Factor)算法，筛选出在局部区域内密度明显不同于其他点的异常值。其计算过程如下：

首先，计算各个数据的可达距离，

$$r(a,b)=\max\left(k_d(a),d(a,b)\right) \tag{6}$$

其中，$r(a,b)$是点a到点b的可达距离；$k_d(a)$是点a的k-距离，即第k个最近邻的距离；$d(a,b)$是点a和b之间的欧氏距离。

其次，引入点a到点b的高斯加权距离：

$$\omega\left(d(a,b),\sigma\right)=\exp\left(-\frac{d(a,b)^2}{2\sigma^2}\right) \tag{7}$$

其中，$d(a,b)$是点α和点b之间距离，σ是标准差。

计算高斯加权的局部可达密度，

$$LRD(a)=\frac{1}{\frac{1}{\left|N_k(\alpha)\right|}\sum_{b\in N_k(\alpha)}r(a,b)\cdot\omega\left(d(a,b),\sigma\right)} \tag{8}$$

其中，a是数据点；$N_k(a)$是点a的k个最近邻。

最后，计算点a的 LOF 值，

$$LOF(a)=\frac{1}{\left|N_k(\alpha)\right|}\sum_{b\in N_k(\alpha)}\frac{LRD(a)}{LRD(b)} \tag{9}$$

其中，$LRD(a)$是点a的局部可达密度；$LRD(b)$是点b的局部可达密度。

3. 基于 GLOF-BFL-LightGBM 的信用评分模型

LightGBM 作为一种基于树模型的集成学习方法，在信用评分领域取得了较好的结果。然而，在样本极度不平衡的情况下，该模型在少数类样本的分类效果仍然存在一定的不足。针对信贷数据不平衡问题，本文将 FocalLoss 损失函数引入到 LightGBM 中，并使用贝叶斯优化 FocalLoss 中的α平衡因子和γ调制因子两个超参数，提升模型对不同信贷样本数据集的泛化能力。针对 FocalLoss 可能过多关注离群点这些错误分类的样本导致 LightGBM 过拟合的问题，本文采用了一种高斯加权调整局部可达密度的 LOF 算法，筛除在局部区域内密度明显不同于其他点的异常值。因此本文提出将高斯加权的 LOF

技术、贝叶斯优化 FocalLoss 损失函数和 LightGBM 算法相结合的继承信用评分模型，模型主要包括三个部分：a) 使用 GLOF 首先筛除信贷样本数据中的极端离群点；b) 使用贝叶斯优化 FocalLoss 的 LightGBM 算法训练数据，直至模型完全收敛；c) 使用训练得到的模型对测试集数据进行分类，得出最终的预测结果。

模型训练过程如下：

输入：初始样本集；

1) 初始化高斯加权的标准差参数 σ 和 LOF 阈值，使用高斯加权调整局部可达密度的 LOF 算法识别并去除样本中的极端离群点，确保训练数据的质量；

2) 使用贝叶斯优化 LightGBM 的 FocalLoss 损失函数超参数，通过多次迭代，搜索超参数空间，找到能够使模型性能最佳的 α 平衡因子和 γ 调制因子超参数配置，得到该样本集最佳超参数；

3) 根据步骤 2)中获得的最佳超参数，重新训练 LightGBM 模型，得到训练后的最终模型；

4) 使用训练好的最终模型对测试集进行预测，得到每个样本的违约概率。

输出：样本违约概率，表示每个样本在模型预测下的违约风险。

4. 实验与结果

4.1. 数据与处理

本文使用了 UCI 数据库中的四个信贷评分数据集：D1、D2、D3 和 D4。有关这四个数据集的详细信息，请参见表 1。

Table 1. Dataset information
表 1. 数据集信息

数据集	特征数	样本数	正负样本比例
D1	14	690	1:1.2498
D2	20	1000	1:2.3333
D3	14	690	1:1.2498
D4	17	6000	1:5.0000

为了客观评估 GLOF-BFL-LightGBM 模型在信用评分领域的性能，本研究在多个公开的信贷数据集上进行了实验。在确保评估的公正性和结果的可复现性的要求下，所有数据集均按照 8:1:1 的比例被随机划分为训练集、验证集和测试集。GLOF-BFL-LightGBM 模型首先利用 80%的训练集进行参数学习和优化，然后通过 10%的验证集对 FocalLoss 进行贝叶斯超参数优化，以确定最优的超参数设置。最后，模型在剩余的 10%测试集上进行评估，从而衡量其在信用评分任务中的泛化能力和实际表现。

4.2. 对比模型及性能评估指标

为全面评估各类信用评分模型的性能与适用性，本文构建了一个系统且深入的模型对比框架，旨在从多个维度细致剖析各模型在信用评分领域中的表现及其潜在优势。基准模型包括几类常见的信用评分方法：统计学方法中的 LDA (线性判别分析)和 LR (逻辑回归)；传统机器学习算法中的 DT (决策树)、KNN (K 近邻)、SVM (支持向量机)和 NN (神经网络)；集成学习方法中的 AdaBoost、GBDT (梯度提升树)、XGBoost；此外，还包括基于不平衡数据采样的集成模型，如 RUSBoost、SMOTEBagging、AsymBoost 和

BalanceCascade。

本文采用六种常用的信用评分评价指标进行模型评估：AUC 值、ACC (准确率)、Prec (精确率)和 Recall (召回率)、F1 分数、GMean 值。其中，AUC 和 ACC 主要用于评估二分类模型的整体表现；Prec 和 Recall 则适用于错判代价较高的情境，在信用评分中，漏判违约客户的成本通常高于误判非违约客户的成本；F1 分数和 Gmean 值则更适合处理类别不平衡问题时的综合评价。

4.3. 实验结果分析

表 2 列出了各个信用评分模型在 D1 数据集上所达到的性能指标。GLOF-BFL-LightGBM 在 AUC、Acc、Recall、F1、GMean 五个指标中都取得了最优，显示出该模型在履约和违约样本的区分能力上具有明显优势，能够全面、准确地识别高风险客户和优质客户。同时，GLOF-BFL-LightGBM 在多个评估指标上的稳定表现也体现了其较强的鲁棒性，适用于在复杂的信贷评分任务中做出可靠的风险评估。尽管 LDA 在 Precision 上表现出较好的成绩，意味着它在减少误判优质客户(即减少假阳性)的风险方面有所优势，但由于其在 AUC、F1 和 GMean 等关键指标上的表现较差，导致其无法全面衡量信贷风险。

Table 2. Performance comparison results on the D1 dataset
表 2. D1 数据集上的性能对比结果

Method	AUC	Acc	Prec	Recall	F1	GMean
LDA	0.91265	0.86058	**0.94020**	0.79969	0.86427	0.86542
LR	0.91560	0.85493	0.91748	0.81164	0.86132	0.85891
DT	0.91344	0.84901	0.86981	0.85614	0.86292	0.84596
KNN	0.91109	0.84867	0.88621	0.83452	0.85959	0.85027
SVM	0.86821	0.85658	0.93115	0.80084	0.86109	0.86108
NN	0.91773	0.84868	0.88950	0.83102	0.85926	0.85032
AdaBoost	0.92105	0.85484	0.92933	0.79927	0.85940	0.85945
GBDT	0.93621	0.86415	0.88977	0.86245	0.87590	0.86436
XGBoost	0.93621	0.86784	0.89586	0.86251	0.87887	0.86849
RUSBoost	0.90307	0.81884	0.83784	0.82667	0.83221	0.81805
SMOTEBagging	0.92899	0.85507	0.87671	0.85333	0.86486	0.85524
AsymBoost	0.91503	0.86957	0.91304	0.84000	0.87500	0.87178
BalanceCascade	0.94596	0.87440	0.93069	0.83186	0.87850	0.87745
GLOF-BFL-LightGBM	**0.96558**	**0.89706**	0.92857	**0.90698**	**0.91765**	**0.89339**

表 3 列出了各个信用评分模型在 D2 数据集上所达到的性能指标。GLOF-BFL-LightGBM 在 AUC、Acc、Precision 和 F1 等关键指标上均取得了最优结果，显示出其在信贷样本好坏识别方面的优异能力，能够有效区分履约客户与违约客户。尽管 AdaBoost 和 LDA 在 Recall 和 GMean 指标上相对较强，能够在一些情形下更好地捕捉违约客户(即高风险客户)，但由于这些模型在其他关键评估指标(如 Precision 和 F1 等)上的表现相对较弱，因此无法全面、准确地评估信用风险。在信贷评分中，尤其是在对风险进行全面管控的场景下，模型的综合表现尤为重要，而 GLOF-BFL-LightGBM 在多个指标上的平衡能力，使其在各种情境下都能够保持较高的评分精度与稳定性。

Table 3. Performance comparison results on the D2 dataset
表 3. D2 数据集上的性能对比结果

Method	AUC	Acc	Prec	Recall	F1	GMean
LDA	0.77954	0.75846	0.79256	0.88714	0.83719	0.63756
LR	0.78084	0.76012	0.79421	0.88720	0.83813	0.64133
DT	0.70962	0.72324	0.77911	0.84389	0.81020	0.59916
KNN	0.73835	0.72798	0.73412	0.95857	0.83146	0.42669
SVM	0.71123	0.70650	0.79660	0.77983	0.78813	0.36659
NN	0.77994	0.76592	0.80674	0.87529	0.83961	0.66778
AdaBoost	0.70350	0.70208	0.70314	**0.99409**	0.82368	0.14356
GBDT	0.77916	0.75870	0.78790	0.89666	0.83877	0.62583
XGBoost	0.78113	0.75816	0.77567	0.92083	0.84204	0.59045
RUSBoost	0.69805	0.69000	0.75000	0.82609	0.78621	0.56549
SMOTEBagging	0.71196	0.71500	0.76821	0.84058	0.80277	0.60503
AsymBoost	0.71996	0.70000	0.76000	0.82609	0.79167	0.58858
BalanceCascade	0.75585	0.74000	0.80788	0.80788	0.80788	**0.69503**
GLOF-BFL-LightGBM	**0.79334**	**0.81000**	**0.80851**	0.98701	**0.88889**	0.46322

表 4 列出了各个信用评分模型在 D3 数据集上所达到的性能指标。GLOF-BFL-LightGBM 在 AUC、Acc、Recall、Precision、F1 和 GMean 等多个评估指标中均取得了最优结果，展现了其在信用评分任务中的卓越性能。同时，Precision 指标仅次于 AsymBoost 和 KNN，进一步证明了 GLOF-BFL-LightGBM 在精准识别违约客户方面的优势。这表明，GLOF-BFL-LightGBM 不仅在提高整体预测准确性方面具有显著优势，还能够在信用评分中有效减少误判风险，有助于金融机构更好地识别高风险客户并优化资源分配。

Table 4. Performance comparison results on the D3 dataset
表 4. D3 数据集上的性能对比结果

Method	AUC	Acc	Prec	Recall	F1	GMean
LDA	0.92688	0.85942	0.79607	0.91961	0.85339	0.86369
LR	0.92977	0.86487	0.83094	0.87414	0.85199	0.86575
DT	0.91402	0.84368	0.82704	0.82020	0.82360	0.83549
KNN	0.91336	0.84939	0.86399	0.78508	0.82265	0.82049
SVM	0.92622	0.86255	0.84973	0.83954	0.84461	0.85257
NN	0.91476	0.85017	0.83282	0.82984	0.83133	0.84778
AdaBoost	0.92733	0.85554	0.79128	0.91726	0.84963	0.85986
GBDT	0.93922	0.86371	0.84260	0.85303	0.84778	0.86260
XGBoost	0.93942	0.86333	0.84487	0.84866	0.84676	0.86178

续表

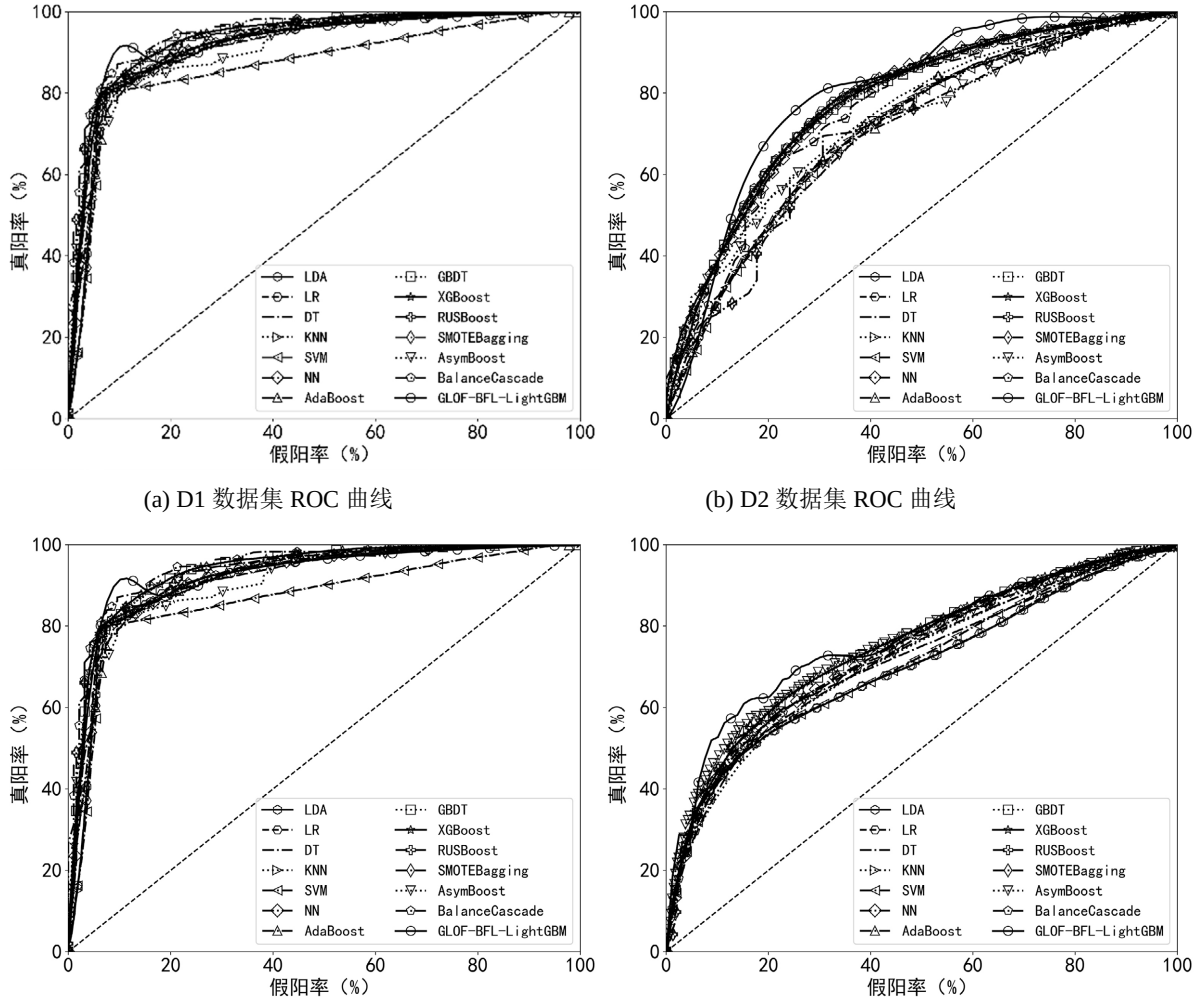
RUSBoost	0.90399	0.86957	0.85106	0.78431	0.81633	0.84924
SMOTEBagging	0.91977	0.85507	0.80392	0.80392	0.80392	0.84351
AsymBoost	0.91120	0.86957	**0.86667**	0.76471	0.81250	0.84378
BalanceCascade	0.92994	0.85990	0.80952	0.83951	0.82424	0.85610
GLOF-BFL-LightGBM	**0.96791**	**0.91304**	0.86111	**0.96875**	**0.91177**	**0.91534**

表 5 列出了各个信用评分模型在 D4 数据集上所达到的性能指标。GLOF-BFL-LightGBM 在 AUC、Acc、Recall、Precision、F1 和 GMean 等多个评估指标中均取得了最优结果，展现了其在信用评分任务中的全面优势。特别是在 AUC 和 Acc 指标上，GLOF-BFL-LightGBM 能够准确地评估样本的整体质量，同时在 Recall 和 Precision 指标上取得平衡，确保在识别违约客户的同时，减少误报的风险。F1 和 GMean 的优秀表现进一步证明了模型在处理不平衡数据集时的稳定性和鲁棒性。

Table 5. Performance comparison results on the D4 dataset
表 5. D4 数据集上的性能对比结果

Method	AUC	Acc	Prec	Recall	F1	GMean
LDA	0.69845	0.65116	0.66763	0.60228	0.63327	0.64933
LR	0.69988	0.64862	0.66120	0.60986	0.63449	0.64747
DT	0.71990	0.66657	0.68511	0.61669	0.64910	0.65889
KNN	0.71690	0.66755	0.71635	0.55495	0.62541	0.65800
SVM	0.70582	0.67314	0.75608	0.51136	0.61010	0.63649
NN	0.73771	0.68058	0.71314	0.60437	0.65426	0.67619
AdaBoost	0.71697	0.67278	0.75990	0.50533	0.60701	0.65163
GBDT	0.74961	0.69477	0.72968	0.61893	0.66976	0.69063
XGBoost	0.75038	0.69453	0.73006	0.61749	0.66907	0.69026
RUSBoost	0.75057	0.70111	0.73427	0.60137	0.66121	0.69146
SMOTEBagging	0.73662	0.67750	0.35338	0.62267	0.65539	0.67453
AsymBoost	0.75221	0.69583	0.74249	0.58545	0.65468	0.68563
BalanceCascade	0.73578	0.67722	0.68342	0.62314	0.65189	0.67360
GLOF-BFL-LightGBM	**0.76665**	**0.71667**	**0.77372**	**0.66250**	**0.71381**	**0.71820**

为了直观展示各信用评分算法的性能，实验部分采用 ROC 曲线进行图形化比较。ROC 曲线是一种常用的评估信用评分效果的图形化工具，其中横轴表示假阳性率(False Positive Rate, FPR)，纵轴表示真阳性率(True Positive Rate, TPR)。图 1 展示了不同信用评分模型在四个数据集上的 ROC 曲线表现。信用评分模型在不同数据集上的效果各不相同，反映了模型在各类特征和样本分布上的适应性。在 D2 和 D3 数据集上，GLOF-BFL-LightGBM 模型的表现显著优于其他模型，表明其在这些数据集上具有较强的预测能力和鲁棒性。而在 D1 和 D4 数据集上，尽管 GLOF-BFL-LightGBM 的优势不如前者明显，但其 ROC 曲线下面积(AUC)依然略高，说明该模型在多种环境下仍能保持较为稳定的性能，适合用于实际的信用评分任务。



(a) D1 数据集 ROC 曲线　(b) D2 数据集 ROC 曲线

(c) D3 数据集 ROC 曲线　(d) D4 数据集的 ROC 曲线

Figure 1. Comparison of ROC curves for the four datasets
图 1. 四个数据集的 ROC 曲线对比图

5. 总结

信用评分作为金融风险管理和决策的关键环节，直接影响金融机构的稳健运营和市场竞争力。然而，信贷数据中普遍存在的类别不平衡现象，即违约样本远少于非违约样本，给信用评分模型的构建带来了挑战。为此，本文提出了一种基于 GLOF-BFL-LightGBM 集成的信用评分模型，旨在有效解决信贷数据类别不平衡问题。该模型集成了高斯加权局部异常因子(GLOF)技术、贝叶斯优化的 Focal Loss 损失函数(BFL)和 LightGBM 算法。其中，GLOF 技术用于过滤极端离群点，避免模型过拟合；BFL 通过贝叶斯优化动态调整损失函数参数，提升模型对少数类样本(违约样本)的识别能力；LightGBM 则作为基分类器，高效地构建信用评分模型。在多个公开信贷数据集上的实验结果表明，GLOF-BFL-LightGBM 模型的性能显著优于传统统计方法、经典机器学习算法以及其他集成学习模型。这验证了 BFL 在处理类别不平衡数据和 GLOF 在离群点过滤方面的有效性，为信用评分领域提供了新的方法和视角。未来研究将探索该框架与其他集成算法的结合，并将其应用于更复杂的信贷数据集和企业风险评估等场景，进一步拓展其应用领域。

参考文献

[1] Zedda, S. (2024) Credit Scoring: Does XGboost Outperform Logistic Regression? A Test on Italian SMEs. *Research in International Business and Finance*, **70**, Article ID: 102397. https://doi.org/10.1016/j.ribaf.2024.102397

[2] Gao, Y., Xiao, H., Zhan, C., Liang, L., Cai, W. and Hu, X. (2023) CATE: Contrastive Augmentation and Tree-Enhanced Embedding for Credit Scoring. *Information Sciences*, **651**, Article ID: 119447. https://doi.org/10.1016/j.ins.2023.119447

[3] Durand, D. (1941) Risk Elements in Consumer Installment Financing. National Bureau of Economic Research, 189-201.

[4] Orgler, Y.E. (1970) A Credit Scoring Model for Commercial Loans. *Journal of Money, Credit and Banking*, **2**, 435-445. https://doi.org/10.2307/1991095

[5] Zhang, R. and Qiu, Z. (2020) Optimizing Hyper-Parameters of Neural Networks with Swarm Intelligence: A Novel Framework for Credit Scoring. *PLOS ONE*, **15**, e0234254. https://doi.org/10.1371/journal.pone.0234254

[6] Zhou, M. (2022) Credit Risk Assessment Modeling Method Based on Fuzzy Integral and SVM. *Mobile Information Systems*, **2022**, Article ID: 3950210. https://doi.org/10.1155/2022/3950210

[7] 王重仁, 韩冬梅. 基于超参数优化和集成学习的互联网信贷个人信用评估[J]. 统计与决策, 2019, 35(1): 87-91.

[8] Liu, W., Fan, H. and Xia, M. (2023) Tree-Based Heterogeneous Cascade Ensemble Model for Credit Scoring. *International Journal of Forecasting*, **39**, 1593-1614. https://doi.org/10.1016/j.ijforecast.2022.07.007

[9] 康海燕, 胡成倩. 基于特征提取和集成学习的个人信用评分方法[J]. 计算机仿真, 2024, 41(1): 311-320.

[10] Khalili, N. and Rastegar, M.A. (2023) Optimal Cost-Sensitive Credit Scoring Using a New Hybrid Performance Metric. *Expert Systems with Applications*, **213**, Article ID: 119232. https://doi.org/10.1016/j.eswa.2022.119232

[11] Wu, Y., Huang, W., Tian, Y., Zhu, Q. and Yu, L. (2022) An Uncertainty-Oriented Cost-Sensitive Credit Scoring Framework with Multi-Objective Feature Selection. *Electronic Commerce Research and Applications*, **53**, Article ID: 101155. https://doi.org/10.1016/j.elerap.2022.101155

[12] Bahnsen, A.C., Aouada, D. and Ottersten, B. (2014) Example-Dependent Cost-Sensitive Logistic Regression for Credit Scoring. 2014 13*th International Conference on Machine Learning and Applications*, Detroit, 3-6 December 2014, 263-269. https://doi.org/10.1109/icmla.2014.48

[13] Shen, F., Zhao, X., Kou, G. and Alsaadi, F.E. (2021) A New Deep Learning Ensemble Credit Risk Evaluation Model with an Improved Synthetic Minority Oversampling Technique. *Applied Soft Computing*, **98**, Article ID: 106852. https://doi.org/10.1016/j.asoc.2020.106852

[14] 邵良杉, 周玉. 一种改进过采样算法在类别不平衡信用评分中的应用[J]. 计算机应用研究, 2019, 36(6): 1683-1687.

[15] 陈启伟, 王伟, 马迪, 等. 基于Ext-GBDT集成的类别不平衡信用评分模型[J]. 计算机应用研究, 2018, 35(2): 421-427.