基于机器学习方法的信用风险监测实证研究

吴方智，王睿斌，张圣凯，王愉鉴，卢靖
浙大宁波理工学院，浙江 宁波 315100

DOI: 10.61369/TACS.2025010028

摘　　要： 本研究构建基于 XGBoost 的信用监测系统，通过正则化计算优化模型性能结合特征工程处理欺诈关联特征，并引入 SMOTE 技术平衡样本分布。提出了 XGBoost、SMOTE 与 SHAP 模型解释融合的风控框架，建立精准欺诈特征体系验证了梯度提升算法在金融安全领域的实践优势。

关 键 词： 欺诈检测；XGBoost；机器学习；特征工程

Empirical Study on Credit Risk Monitoring Based on Machine Learning Methods

Wu Fangzhi, Wang Ruibin, Zhang Shengkai, Wang YuJian, Lu Jing
NingboTech University, Ningbo, Zhejiang 315100

Abstract: In this study, a credit monitoring system based on XGBoost is constructed, the model performance is optimized by regularization computation combined with feature engineering to deal with fraud-related features, and SMOTE technology is introduced to balance the sample distribution. A wind control framework integrating XGBoost, SMOTE and SHAP model interpretation is proposed, and the establishment of an accurate fraud feature system verifies the practical advantages of the gradient boosting algorithm in the field of financial security.

Keywords: fraud detection; XGBoost; machine learning; feature engineering

引言

随着电子支付普及，银行卡欺诈呈现隐蔽化、复杂化趋势。传统基于规则的方法难以应对高维交易数据和非线性欺诈模式，尤其对极少数欺诈交易识别能力不足。针对数据高度不平衡、欺诈特征复杂等核心问题，构建智能风控系统非常有意义。引入模型解释技术增强决策透明度，采用参数调优提升泛化能力，保证检测效率的同时显著提升对隐蔽欺诈行为的捕捉能力，构建动态特征学习框架以适应欺诈模式演变，通过可解释性分析为风控决策提供可操作依据，为金融安全防护提供有效技术路径成为重要研究方向。

一、相关工作

银行卡欺诈检测技术历经规则引擎、统计分析和机器学习等阶段演进。早期规则引擎依赖预定义阈值识别异常交易[1]，虽具可解释性却难以应对新型欺诈[2]。统计方法受限于高维非线性数据处理能力[3]，而传统机器学习模型如逻辑回归、决策树存在过拟合和泛化能力缺陷。集成方法如随机森林通过多树投票机制提升稳定性[4]，SVM 凭借核技巧优化分类边界[5]，但二者对数据不平衡问题敏感[6]。深度学习模型虽能自动提取复杂特征[7]，图神经网络更可挖掘交易关联模式[8]，但其黑箱特性制约金融场景应用[9]。

在众多机器学习方法中，XGBoost 凭借高效的梯度提升框架，通过正则化项控制模型复杂度[10]，在处理高维稀疏金融数据时展现显著优势，国内学者提出的分层抽样策略进一步强化了其对不平衡数据的适应性[11]，其树结构特性天然支持特征重要性评估，结合 SHAP 值可量化特征贡献度[12]。实践表明该方法可有效识别交易时空模式等关键风险因子[13]。

随着技术的不断进步，新的研究正在探索更先进的模型和方法。例如，最近的研究提出了将 LSTM 与 XGBoost 相结合的混合模型，以进一步提升模型的性能[14]。总而言之，从传统方法到机器学习，再到深度学习和集成方法，欺诈检测技术不断演进，XGBoost 和 SHAP 值等技术的引入为金融风控领域提供了更高效、更灵活的解决方案[5]。

二、基于 xgboost 机器学习方法的信用风险监测系统

本文基于 xgboost 机器学习方法的信用风险监测系统，主要包括原始数据分析、数据清洗、特征工程、模型构建、模型评估、模型解释等6个流程。下面结合欺诈检测数据集介绍上述流程所采用的方法和理论依据。

（一）数据集

本文使用 Kaggle 的 Credit Card Fraud Detection 数据集，包含六个核心数据表。数据整合为三个维度：客户基础信息、信用行为特征和外部环境指标。客户基础信息包括唯一标识码、性别、年龄等；信用行为特征有贷款金额、收入总额、信用评分等；外部环境指标有区域风险评级、人口密度等。研究旨在通过多维数据分析，构建风险评估体系，支持差异化授信决策。

（二）数据分析

在构建基于 xgboost 机器学习方法的信用风险监测系统之前，通过对原始数据的分布、缺失值以及特征之间的多重共线性进行分析，可以更好地理解数据的特性。

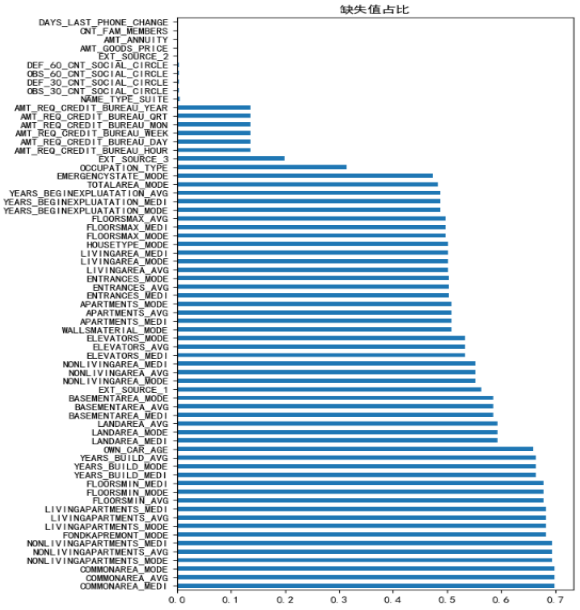

图 1 贷款金额特征值分布图

1. 特征值分布

数据分布图揭示了特征值的密度分布，包括峰值和偏态。分析这些分布有助于理解特征贡献并识别异常值。我们发现多数特征呈正态分布，如图 1 所示，贷款金额分布较低且符合正态分布。

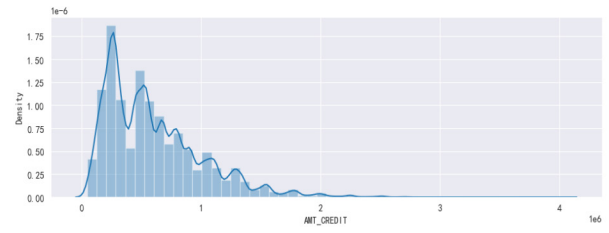

图 2 数据集缺失值占比分布图

2. 缺失值分析

在原始数据中，部分字段存在缺失值。为了确保模型的准确性，我们对缺失值进行了详细分析，并计算了每个字段的缺失值占比。由图 2 所示，数据集中存在大量缺失的情况，这意味着在数据清洗时我们需要慎重的选择处理缺失值的方式。

（三）数据清洗

在处理上述原始数据之前，首先要进行数据清洗，以提高模型的准确性。本文数据清洗方法主要包括缺失值处理以及异常值处理。

1. 缺失值处理

针对 3.2.2 节缺失值分析结果，本文采取差异化处理策略。数值型特征（如交易金额、收入水平）采用迭代回归填补法（IterativeImputer），通过多变量相关性建模实现缺失值预测，有效保持数据维度与变量间关联性。类别型特征（如交易地点、类型）则采用众数填补，避免引入异常值的同时保留原始分布特征。该方法兼顾数据特征类型与业务逻辑，在消除缺失干扰与维护数据结构间取得平衡。

2. 异常值处理

异常值可能源于录入错误或极端情况，影响模型训练。使用 IQR 方法检测异常值，即通过计算上四分位数与下四分位数的差值来识别超出一定范围的数据点。对于数值型字段的异常值，用中位数替换；非数值型字段则用众数替换。极端异常值则直接删除，以维护模型训练的稳定性。

（四）特征工程模块

信贷本身涉及环节众多，影响信贷风险因素十分复杂，各类特征重要性各有差别，因此对特征根据情况进行处理十分必要。

1. 离散化特征的处理

本研究针对类别型特征（如交易类型、商户类别）实施离散化处理，解决模型输入兼容性问题。采用标签编码策略适配树形模型特性，在保留变量间序贯关系的同时避免独热编码引发的维度膨胀与信息损耗。相较于传统编码方法，该方案在增强特征表达能力的基础上，显著提升树模型（如 XGBoost）对欺诈模式的捕获效率，为风险识别提供高鲁棒性数据表征。

2. 处理不平衡数据的方法

在银行卡欺诈检测中，数据不平衡问题可通过三种方法解决：SMOTE 生成合成样本增加少数类，过采样重复少数类样本，欠采样减少多数类样本。SMOTE 是常用的方法，可平衡数据同时避免过拟合。本文采用 SMOTE 方法处理数据不平衡。

（五）模型构建

本文模型构建包括模型参数调优、模型训练、交叉验证，在参数调优前后共进行了两次模型训练。

1.XGBoost 模型的初始化

本研究使用 XGBoost 梯度提升框架，通过 XGBClassifier() 初始化模型，该模型采用经过验证的默认参数，包括自适应学习机制、鲁棒性正则化和动态特征交互。这些配置无需人工干预，即可捕捉信用违约风险的关键模式。

2. 模型训练

XGBoost（极限梯度提升）是一种基于梯度提升决策树的集成学习框架，其训练过程通过多轮次叠加弱学习器（CART 回归树）逐步优化预测结果。该算法在金融风险建模中的应用价值，源于其对高维稀疏特征的非线性映射能力及分布式计算的工程优化。本节从数学角度描述其训练机制：

给定训练集 $D=\{x_i,y_i\}_{i=1}^{n}$，其中 $x_i \in R^d$ 为特征向量，$y_i \in \{0,1\}$ 为二分类标签。模型通过 K 棵回归树构建复合预测函数，其定义如式 (1) 所示：

$$\hat{y}_i = \sigma(\sum_{k=1}^{K} \alpha_k \cdot g_k(x_i;\Theta_k)) \quad (1)$$

式中 $\sigma(z)=(1+e^{-z})^{-1}$ 为概率映射函数，$g_k(\cdot)$ 表示第 k 棵树的输出值，Θ_k 包含树结构参数（分裂特征、叶节点权重等），α_k 为收缩率超参数，用于抑制单棵树的影响。

模型训练最小化以下复合损失函数，如式 (2) 所示：

$$\ell = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(\Theta_k) \quad (2)$$

其中 $\ell(\cdot)$ 采用对数损失函数，如式 (3) 所示：

$$\ell(y,\hat{y}) = -y\ln\hat{y} - (1-y)\ln(1-\hat{y}) \quad (3)$$

正则化项 $\Omega(\Theta_k) = \beta T_k + \frac{1}{2}\mu\|\omega_k\|^2$，$T_k$ 为叶节点数，ω_k 为叶节点权重向量，β，μ 分别控制结构复杂度和权重幅值。

在第 t 次迭代时，基于前 t—1 棵树的预测结果 $\hat{y}_i^{(t-1)}$，通过二阶泰勒展开近似目标函数，目标函数如式 (4) 所示：

$$\ell^{(t)} \approx \sum_{i=1}^{n}\left[\nabla_{\hat{y}^{(t-1)}}\ell(y_i,\hat{y}_i^{(t-1)})\cdot f_t(x_i) + \frac{1}{2}\nabla^2_{\hat{y}^{(t-1)}}\ell(y_i,\hat{y}_i^{(t-1)})\cdot f_t^2(x_i)\right] + \Omega(\Theta_t) \quad (4)$$

令 $g_i = \nabla_{\hat{y}^{(t-1)}}\ell$，$h_i = \nabla^2_{\hat{y}^{(t-1)}}\ell$，则节点分裂增益准则如式 (5) 所示：

$$g_{split} = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \mu} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \mu} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \mu}\right] - \beta \quad (5)$$

其中 I_L，I_R 为分裂后的左右子节点样本集，μ 项实现 L2 正则化。算法通过预排序（pre-sorted）策略和特征分桶（bucketing）加速最优分裂点搜索。

训练终止后，最终预测值由所有树的加权输出决定，其公式如式 (6) 所示：

$$p_i = \sigma\left(\sum_{k=1}^{K} \alpha \cdot g_k(x_i)\right) \quad (6)$$

其中 α 为全局学习率，与 α_k 共同作用形成双重收缩机制。此过程在保证模型泛化能力的同时，显式建模了特征间的高阶非线性交互效应，为信用风险评估提供可解释的决策依据。

3. 模型优化

本文通过系统化网格搜索策略对 XGBoost 模型的关键超参数进行优化，以实现模型泛化能力与计算效率的最佳平衡。核心参数搜索包括学习率 [0.01, 0.1, 0.2]、树最大深度 {3, 5, 7} 和基学习器数量 {50, 100, 200}，以适应不同的数据特性和模型需求。

4. 模型性能评估指标

采用多维度评估体系，重点选取 ROC-AUC 指标衡量模型在不平衡数据中的阈值鲁棒性，同步结合准确率、精确率、召回率及 F1 评分构建四维评估矩阵。AUC 通过真 / 假阳性率动态关系综合评价分类性能，准确率反映整体判别精度，精确率 – 召回率双指标分别控制误报与漏报风险，F1 评分则平衡二者矛盾。该指标体系从全局判别、代价敏感、风险管控等维度形成闭环评估，为欺诈检测模型的迭代优化提供多视角分析框架。

（六）模型解释

为解构 XGBoost 模型在信用卡异常交易识别中的决策逻辑，本研究采用 Shapley 加性解释框架（SHAP）进行特征归因分析。基于 Lloyd-Shapley 值理论，该方法将模型预测值解构为各特征变量的边际贡献组合，其数学表达满足线性可加性公理：

$$\phi_0 + \sum_{j=1}^{m} \phi_j(x_j^{(i)}) = f(x^{(i)}) \quad (11)$$

式中 $f(x^{(i)})$ 表示样本 $x^{(i)}$ 的模型预测输出，$\phi_j \in R$ 为第 j 维特征的 Shapley 值，ϕ_0 表示全体样本的期望预测基准值。

在银行信用卡欺诈检测任务中，SHAP 值方法能够帮助识别对欺诈检测最重要的特征，并量化这些特征对模型预测的具体影响。例如，通过分析 SHAP 值，可以发现高额交易的交易对欺诈检测的正向贡献较大，而低额交易则对欺诈检测的负向贡献较大。

三、实验与结果分析

本研究进行了两次模型训练，第一次使用 122 个特征工程后的特征，结合 SHAP 值筛选最优模型。第二次训练前通过网格搜索确定最优超参数，并进行二次训练以评估其效果。SHAP 方法用于解释实验结果，并通过对比实验验证了 XGBoost 模型的优越性。

（一）实验数据集

本系统使用 Kaggle 的 Credit Card Fraud Detection 数据集，包含 307,511 条交易记录和 122 个特征，其中正常交易 282,686 条，欺诈交易 24,825 条，存在数据不平衡。数据分为申请人信息、通讯信息、贷款信息和房屋信息等类别。采用 imblearn 的 SMOTE 进行过采样，并使用 sklearn 的 train_test_split 划分数据集，训练集和测试集比例分别为 80% 和 20%。

（二）实验环境

实验硬件平台处理器为 Intel Core i7-11800H @ 2.30GHz，内存为 32GB DDR4 3200MHz，固态硬盘容量为 1TB NVMe SSD，GPU 显卡为 NVIDIA GeForce RTX 3060 Laptop GPU。操作系统使用 Windows 11 22H2，软件开发平台为 Jupyter Notebook 6.5.4。

（三）模型性能评估

本节对比了 XGBoost 与其他三种主流机器学习模型（逻辑回归、随机森林、支持向量机）在信用卡欺诈检测的性能。如图 3 显示，XGBoost 在精确度、召回率、F1 值和准确率四项指标上均优于或接近其他模型，其中精确度达 99.56%，召回率为 96.25%，F1 值为 97.88%，准确率为 97.05%，证明了 XGBoost 在实际应用中的优越性和有效性。

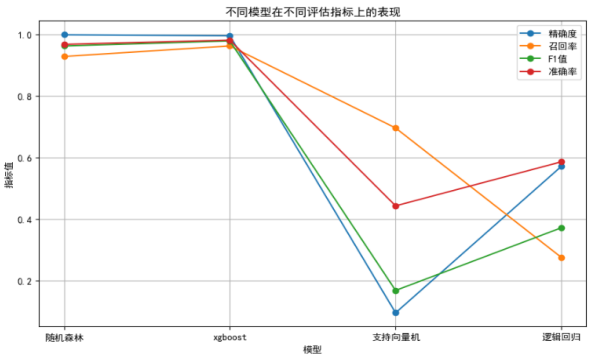

图 3 多种模型的性能对比图

（四）超参数优化结果

在4.3的模型对比中，xgboost 相较于其他模型在评估指标上都有不同幅度的领先。因此超参数优化以 xgboost 为基础，按照3.5.3节所述方法对模型进行参数优化和调整，得知 'learning_rate': 0.2, 'max_depth': 7, 'n_estimators': 200的超参数组合最优，将其带入 xgboost 模型并完成训练，得到性能评估结果，如图4所示。

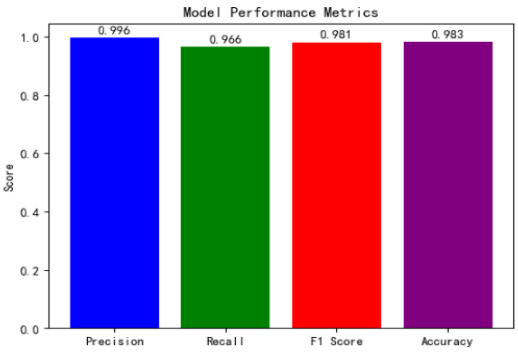

图4 网格搜索优化后的 xgboost 性能指标展示图

相较于未优化的 xgboost 模型，各项评估指标上都有些微提升。

（五）模型实验结果解释

鉴于 XGBoost 模型表现最优，本文采用 TreeSHAP 对其结果进行解释。TreeSHAP 利用 Shapley 值计算特征对预测的贡献，确保特征贡献总和等于模型输出。文中分析了特征对预测的贡献、单个特征的影响、特征间关系及模型决策原因。

1. 各项特征对模型预测结果的贡献解释

由于 SHAP 值是针对某一条记录中的某个特征对于该模型的贡献，因此如果要观察该特征的总体 SHAP 值情况来查看其对模型的贡献，就需要收集每一条记录中该特征对应的 SHAP 值，最终通过其平均数来展示。本文使用 shap.summary_plot 函数计算各个特征的 SHAP 值，并展示贡献较高的前20个特征，排序如图5所示。

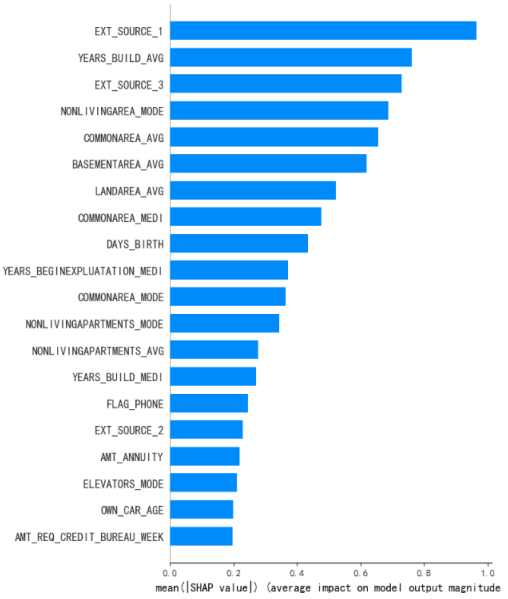

图5 特征全局 SHAP 值排序图

图5的20个特征解释见图6，其展示了特征的全局影响。图表通过 SHAP 值展示了每个特征对预测的影响程度和方向，特征名在纵轴，SHAP 值在横轴。正 SHAP 值表示特征增加欺诈概率，负值则相反。

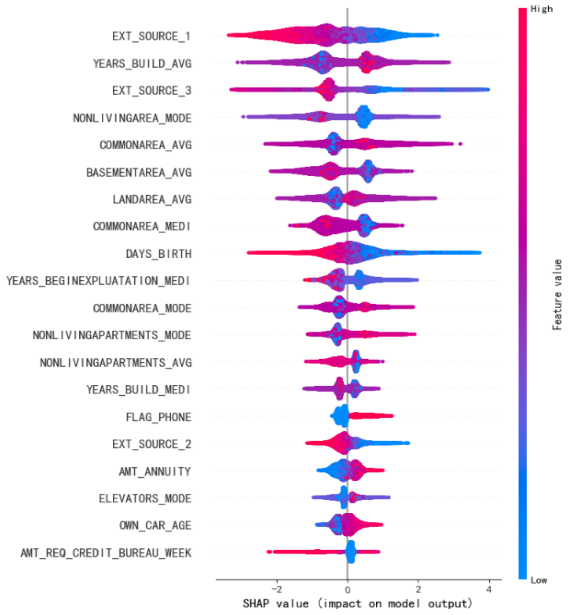

图6 特征全局解释图

如图6所示，不同颜色的点代表了特征取值的高低，红色点表示特征值较高，蓝色点则表示特征值较低。从图中可以明显看出，EXT_SOURCE_1该特征对信用欺诈风险的影响最为显著，颜色为蓝色的样本点普遍位于右半图，也即是 SHAP 值高于0，颜色为红色的样本点普遍位于左半图，也就是 SHAP 值低于0，表明贷款人的外部数据源分数越高，其欺诈的可能性就越小。其次是 YEARS_BUILD_AVG，这一特征的值越小，信用欺诈的风险就越低。这些发现与行业内的实践经验相吻合。

2. 模型决策原因解释

为帮助信贷风险评估部门理解模型对欺诈样本的决策，我们用 SHAP 值分析了高风险样本。图7展示了单个欺诈记录的 SHAP FORCE PLOT，揭示了影响预测的特征及其相对影响大小。

图7 模型决策解释图

决策排序根据样本预测结果的重要性降序排列，SHAP 值显示特征对预测的贡献。红色箭头表示增加欺诈概率的特征，蓝色箭头则相反。例如，建筑平均年限和外部信用评分对预测有显著负影响，表明资产稳定性和信用评分在风险识别中的作用。土地和公共区域面积则显示正向驱动，暗示地理特征与欺诈行为的联系。建议将土地特征纳入风险评估，研究公共区域与资金流动性的关系。

四、结论与未来工作

本文构建了基于 XGBoost 的信用风险监测系统，以解决金融欺诈检测中的适应性和解释性问题。实验显示，该模型在精确率和召回率上优于传统方法，其自适应机制和树结构优化提升了捕捉时序模式的能力。通过 SHAP 框架，建立了预测与特征的量化关系，形成了预测与解释相结合的决策体系。分析验证了模型与金融知识的契合度，揭示了非线性交互效应，为识别新型欺诈策略提供了优化空间，证实了可解释机器学习在金融场景中的应用价值。

参考文献

[1] 张志强，李志强 . 统计欺诈检测方法研究 [J]. 统计科学，2002.17(3),235-249.

[2] Phua C. 李 ,V, K, 与盖勒 ,r . 史密斯 . 基于数据挖掘的欺诈检测研究综述 [J].2010.arXiv 预印本 arXiv:1009.6119.

[3] 韩德杰，韩文杰 . 消费信用评分的统计分类方法研究 [J]. 皇家统计学会杂志：系列 A（社会统计），1997.160(3),523-541.

[4] 李建民，李建民，李建民 . 随机森林 [J]. 机器学习，2001.45(1),5-32.

[5] 陈志强，陈志强 . 支持向量网络 [J]. 机器学习，2006.20(3),273-297.

[6] 戴尔波佐洛，A., Caelen， O., Le Borgne， Y. A., waterschot， S.，和 Bontempi， G. 基于实证分析的信用卡欺诈检测方法 [J]. 专家系统与应用，2014.41(10),4915-4928.

[7] 李昆，Y., Bengio， Y.，和 Hinton， G. 深度学习 [J]. 自然，2015.521(7553),436-444.

[8] 王丹，刘强 . 基于图像神经网络的欺诈交易检测 [J]. 第 27 届国际人工智能联合会议论文集，2018.3940-3946.

[9] 李建军，李建军，李建军，等 . 可解释性机器学习：一种可解释的黑箱模型 [J].Leanpub.

[10] 陈涛，Guestrin， C. 一种可扩展的树助推系统 [J]. 第 22 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集，2016.785-794.

[11] 刘建军，刘建军，刘建军，等 . 数据挖掘技术在信用卡欺诈中的应用 [J]. 决策支持系统，2008,30(3),662-663.

[12] 李淑娟，李晓明，李晓明 . 气候变化对气候变化的影响 [J]. 神经信息处理系统的进展，2017.4765-4774.

[13] 李建军，李建军，李建军，等 . 可解释性机器学习：一种可解释的黑箱模型 [J].Leanpub.

[14] 傅友 , 邹东升 .SDN 中基于条件熵和决策树的 DDoS 攻击检测方法 [J]. 重庆大学学报 ,2023,46(07):1-8.

[15] 张静 . 基于家庭特征的个人信用风险评估研究 [D]. 贵州大学 ,2024.DOI:10.27047/d.cnki.ggudu.2024.001048.

