



基于LLM-RFE-XGBoost方法的客户升级投诉预警研究

乔雅倩¹, 王倩¹, 刘芳¹, 徐宇¹, 江君²

(1. 中移在线服务有限公司山东分公司, 山东 济南 250022;

2. 中移在线服务有限公司AI能力中心, 河南 郑州 450001)

摘要: 客户升级投诉作为衡量服务质量的关键指标, 其精准预测对服务质量改善与客户问题解决具有重要意义。基于此提出了一种基于大语言模型 (LLM)、递归特征消除 (RFE) 以及XGBoost的混合学习方法对潜在升级投诉客户进行预警, 首先利用大语言模型从客户通话文本中提取语义特征, 再融合结构化数据后利用RFE进行最优特征选择, 最后基于有效特征采用XGBoost进行预警。为了检验该模型的有效性, 以某省运营商生产数据为研究对象进行了预测, 实证结果表明提出的LLM-RFE-XGBoost混合方法具有最优的预测性能。实际上线应用后, 升级投诉量下降了6.7%, 成效显著, 对提高服务质量及客户满意度具有重要意义。

关键词: 升级投诉; XGBoost; 递归特征消除; 大语言模型

中图分类号: TP391.4, TP181

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2026033

Research on customer escalated complaints prediction based on LLM-RFE-XGBoost approach

Qiao Yaqian¹, Wang Qian¹, Liu Fang¹, Xu Yu¹, Jiang Jun²

1. China Mobile Online Service Co., Ltd., Shandong Branch, Jinan 250022, China

2. China Mobile Online Service Co., Ltd., AI Capability Center, Zhengzhou 450001, China

Abstract: Customer escalated complaints serve as a critical indicator for measuring service quality and are vital for improving service quality and resolving customer issues. A hybrid learning approach LLM-RFE-XGBoost was proposed for early warning of potential escalated complaints. Firstly, large language models (LLM) was utilized to extract semantic features from customer call text. Then these features were integrated with original structured data, after which recursive feature elimination (RFE) was applied to select the optimal feature set. Finally, XGBoost was employed for prediction using all selected features. To validate the effectiveness of the model, predictive analysis was conducted using production data from a provincial telecom operator as the research subject. Empirical results demonstrate that the proposed LLM-RFE-XGBoost hybrid approach delivers optimal predictive performance. After practical application in a provincial telecom operator, escalated complaints decreased by 6.7%, which is of great significance for improvement of the service quality and customer satisfaction.

Key words: escalated complaint, XGBoost, recursive feature elimination, large language model

收稿日期: 2025-07-16; 修回日期: 2025-11-20

通信作者: 乔雅倩, qiyaqi_06@163.com

0 引言

随着移动业务的发展与客户需求量的不断增长,投诉问题也在不断增加,有效解决客户问题、提升客户满意度成为运营商面临的重要挑战。作为服务质量的关键考核指标,客户升级及有责投诉引发了广泛关注。对有升级倾向的投诉客户进行预判、前置识别、提前安抚、匹配高技能座席等前置防控工作,可为相关工作人员提供有价值的参考,针对可解决的问题进行优先解决,是企业维系声誉、保障发展以及提升客户满意度的重要举措。本文旨在通过智能算法探索升级投诉客户特征并进行及时预警,以期为服务质量提升提供有意义的参考。

近些年来,学者们对与客户投诉相关的问题进行了诸多研究,关于客户投诉与流失的分析与预警方法,主要可分为2类,分别为经典的统计学方法和机器学习模型。统计学方法主要包括聚类分析、Logistic回归等^[1-3]。例如,李青等^[4]利用聚类分析方法和决策树探索影响客户投诉的主要原因,并据此对投诉客户进行预测,通过建立预警体系减少客户投诉与流失;张艳丽^[5]通过Logistic回归构建了顾客抱怨管理机制,通过回归方法与假设检验,对顾客特性因素和情境因素进行分析。客户升级投诉行为受到多种可变因素的影响,具有较强的非线性和复杂性,这类模型在处理非线性数据时,通常不能达到令人满意的效果。随着数据挖掘技术的兴起,机器学习方法以其优越的预测性能受到了广泛关注,被广泛应用于客户投诉预警中,主要包括人工神经网络(artificial neural network, ANN)以及以决策树为基础的各类集成模型等^[6-10],马晓亮等^[11]建立了人工神经网络模型,并结合逻辑回归对运营商客户重复投诉进行分析与预测;李龙戈等^[12]提出基于森林元学习网络(ensemble forest meta-learning network, EFML)的客户流失预警思路,首先构建

客户时序特征的语义向量,再通过集成多个基础树模型生成最终预测结果。实验证明,相比于随机森林(random forest, RF),EFML可以显著提升模型预测精度。张爱华等^[13]从客户投诉文本入手,利用Word2vec对文本进行向量化与相似度计算,并通过隐含狄利克雷分布(latent Dirichlet allocation, LDA)对投诉文本主题进行提取,总结出了客户升级投诉五大主要原因。穆晓玛^[14]利用BP神经网络、时间序列以及多元线性回归模型对客户投诉工单量进行预测,实证结果表明BP神经网络具有最优的预测性能。机器学习方法在结构化数据处理预测方面有着较好表现,但对于非结构化数据的语义解析和情感识别仍存在一定的局限性,而客户情感本身具有重要价值^[15]。以往研究大多聚焦于客户行为、基础信息等结构化数据,或通话文本、工单内容这类非结构化数据进行分析,面对众多复杂的业务规则与海量用户通话文本,单一模型表征能力受限,影响预测准确性的持续提升。

尽管大语言模型(LLM)在文本表征领域展现出显著优势,但其在客户升级投诉预警中的应用尚未得到充分探索。为此,本文提出一种基于大语言模型与轻量机器学习模型融合的预警框架(LLM-RFE-XGBoost),首先利用大模型从非结构化文本数据提取结构化特征,然后融合原有特征后利用递归特征消除进行特征选择,最后利用预测性能较优的XGBoost进行分类。混合学习方法可以有效降低过拟合,提升模型的泛化性能^[16]。该框架通过本地化部署大模型,实现数据闭环处理,有效规避隐私泄露风险。主要贡献如下。

(1) 非结构化数据深度提取:利用大模型对投诉通话文本以及工单内容进行深度语义提取,捕捉其情感倾向以及核心诉求、升级概率等多维结构化特征。

(2) 结构化特征动态选择:针对客户历史行为等结构化数据,利用基于随机森林的递归特征



消除 (RFE) 算法, 筛选对客户升级行为有重要影响的关键指标。

(3) 双模型协同预警框架: 本文创新性地提出并验证了大小模型融合预警方法 (LLM-RFE-XGBoost), 该方法将 LLM 输出结构化特征拼接后输入 XGBoost 分类器, 对潜在升级投诉客户进行预警, 旨在实现精确率与低资源消耗的平衡。根据某省级运营商升级投诉数据进行实证分析, 结果证明 LLM-RFE-XGBoost 预警方法在升级投诉客户分类中相较于单一模型具有较高的预测性能。

1 方法

1.1 基于 XGBoost 的预测算法

XGBoost 是一种基于 Boosting 的集成算法, 由陈天奇^[17]于 2016 年提出, 凭借其快速、高准确率等优势被广泛用于分类和回归等任务中。该算法在每次迭代中, 会根据当前模型预测结果与实际结果的误差来训练下一个模型, 最后通过加法模型将若干基学习器集成进行预测。目标函数由经验风险和结构风险组成, 经验风险是损失函数的累加, 而结构风险由正则化项进行表示。这种设计可以有效平衡模型性能与复杂度。XGBoost 最大的优点在于对损失函数进行了二阶泰勒展开, 二阶导数信息可以精确捕捉损失曲面曲率, 使其优化方向更接近全局最优解, 增强了优化效率与泛化性能。目标函数表示如下:

$$\text{Loss} = \sum_{i=0}^n l(y_i, \hat{y}_i) + \sum_{i=0}^n \Omega(f_i) \quad (1)$$

$$\Omega(f_i) = \sum_{i=0}^n \gamma T + \lambda \frac{1}{2} \sum_{j=0}^T \omega_j^2 \quad (2)$$

其中, $l(y_i, \hat{y}_i)$ 用来衡量预测值和真实值之间的差距, 表征模型拟合能力, 在分类问题中通常使用交叉熵损失函数表示; $\Omega(f_i)$ 为正则化项, T 表示叶子节点的数量, γ 为超参数控制正则化的程度, 提升模型泛化性能, 避免产生过拟合。当正则化参数为零时, XGBoost 即传统的 Boosting 模型。

XGBoost 模型中, 采用分类与回归树 (clas-

sification and regression trees, CART) 作为其基学习器。在模型训练过程中, 为寻找最优的树结构, 算法通过迭代评估候选节点分裂方案, 计算由分裂带来的增益损失并选择增益损失最大的节点进行分裂。对于分类任务, 节点分裂时常采用基尼系数 (gini index) 作为衡量节点不纯度的指标, 其计算式如下:

$$\text{Gini} = \sum_{k=1}^K [p_k(1-p_k)] = 1 - \sum_{k=1}^K p_k^2 \quad (3)$$

其中, p_k 表示类别样本的比例。

1.2 递归特征消除

基于随机森林的递归特征消除 (recursive feature elimination, RFE) 是一种嵌入式特征选择方法, 核心机制在于迭代训练随机森林模型, 依据特征重要性度量对特征进行排序, 并递归剔除排序末位的特征子集, 最终收敛于高判别性特征的最优子空间。本文利用基尼不纯度减少量计算特征重要性, 采用交叉验证的方式对不同特征集进行验证, 据此确定并保留对目标变量贡献最大的特征集。该方法通过找到一个在交叉验证下表现最优的特征子集减少模型过拟合的风险, 其递归消除的步骤如下:

(1) 评估模型与交叉验证策略选择, 本文采用的是随机森林与 K 折交叉验证策略。

(2) 模型训练与特征评估, 利用随机森林针对所有特征进行训练, 输出每个特征的重要性得分。

(3) 特征剔除与交叉验证, 剔除重要性排序得分最小的特征, 利用剩余特征在交叉验证集上重新训练模型。

(4) 迭代剔除, 重复特征剔除与模型训练过程, 直到达到指定特征数量或无法继续剔除, 确定最优特征子集。

1.3 大语言模型

大语言模型 (large language model, LLM) 是基于 Transformer 架构的自回归模型, 其核心是通

过海量文本数据训练具备千亿级参数规模的深度神经网络模型，能够理解和生成类人文本。系统通过 API 调用的方式接收用户端请求并转发至底层 LLM 的 API 服务，将任务指令 (prompt) 传递至模型推理引擎。prompt 在此过程中充当显式控制信号，通过其语义结构引导模型生成与预期任务严格对齐的输出。通过系统化的指令工程 (prompt engineering) 对 prompt 进行形式化优化，其中包含上下文约束注入、角色设定及输出格式声明，可显著提升模型在复杂任务中的行为可控性与输出准确性，最终实现从开放域对话到领域特定任务的高效适配。

1.3.1 LLM 特征提取设计

本文选用 DeepSeek 基于 Qwen-32B 蒸馏得到的 DeepSeek-R1-Distill-Qwen-32B 模型，该模型是 DeepSeek 在推理优化方向上的重要成果，综合性能优异且适合本地化部署，可以有效避免敏感数据泄露。通过指令微调的方式，利用 DeepSeek-R1-Distill-Qwen-32B 从通话文本与工单内容中提取结构化特征。除了通话文本外，额外引入自然语言任务指令与输出示例。prompt 描述如下：

您作为通信运营商行业的优秀质检员，任务是通过分析用户的通话文本与工单内容，完成以下任务。

(1) 情感得分：输出愤怒-平静-高兴的强度评分[-1, 1]；

(2) 关键诉求：提及退款、加急处理等核心诉求的次数；

(3) 负面行为：统计含有威胁性及攻击性语言（如“我要向工业和信息化部（简称工信部）投诉”）的次数。

(4) 升级意图：判断升级投诉概率，范围为[0,1]。

输出格式：以字典格式输出

包含字段：emotion, demands, threats, complaint

例如：{'emotion': ", 'demands': ", 'threats': ", 'complaint':"} }

1.3.2 LLM 有效性验证

LLM 提取特征示例见表 1，以客户实际来电语音文本为例，展示了 LLM 的特征提取数据。以高风险客户 A 为例，LLM 识别出其负面情绪、要退款的关键诉求、工信部投诉的负面行为以及较高的升级投诉意图，LLM 提取的文本特征与人工理解基本一致。

为了保证 LLM 提取的 4 个文本特征指标真实可用，随机选取了 100 条通话文本，由 2 名具备客服经验的员工按照规则进行标注。基于 100 条人工标注文本，利用平均绝对误差 (mean absolute error, MAE) 对 LLM 提取的指标误差进行量化。平均绝对误差定义为式 (4)。人工标注与 LLM 提取特征误差见表 2，情感得分的误差为 0.15；关键诉求误差为 0.21，表明 100 条标注文本中 LLM 提取的客户关键诉求次数与人工标注次数的平均误差为 0.21 次；出现负面行为次数误差为 0.14 次，升级意图得分误差为 0.19。可以看出，LLM 提取的情感得分、关键诉求、负面行为与升级意图与人工标注结果一致性较高，可用于

表 1 LLM 提取特征示例

| 客户简化文本 | 情感得分 | 关键诉求 | 负面行为 | 升级意图 |
|--|------|------|------|------|
| 客户 A: 这个月流量扣了我 30 元，为什么只退 15 元，这不是诈骗吗，我要去工信部投诉，把费用都给我退回来 | -0.4 | 1 | 1 | 0.8 |
| 客户 B: 家里宽带没信号啊，已经打过好几次电话了一直没有解决，你给你们上级反映吧 | 0.1 | 0 | 1 | 0.1 |
| 客户 C: 我看流量还有 800 MB，过了一会就发短信欠费了，怎么延迟那么久，解决不了就抓紧给我上报退回来 | -0.2 | 2 | 1 | 0.6 |
| 客户 D: 问题是你们没及时加上流量，扣了 3 元的流量费，本来就是你们的问题，你必须把这个费用给我退回来 | -0.6 | 1 | 0 | 0.5 |



客户通话文本的特征提取。

$$MAE = \sum_{i=1}^n |L_i - M_i| \quad (4)$$

其中, L 表示 LLM 的输出结果, M 表示人工标注结果。

表2 人工标注与 LLM 提取特征误差

| 特征类型 | 情感得分 | 关键诉求 | 负面行为 | 升级意图 |
|-------|------|------|------|------|
| MAE 值 | 0.15 | 0.21 | 0.14 | 0.19 |

1.4 LLM-RFE-XGBoost 混合学习方法

为了提高模型预测的精度, 对潜在升级投诉客户进行有效且精准的预测, 本文提出了一种基于大小模型融合的 LLM-RFE-XGBoost 混合学习框架, 主要由 3 个部分构成。

第 1 部分: LLM 特征提取。利用 LLM 对文本数据中的情感得分、关键诉求、负面行为以及升级意图进行提取, 并将其转化为结构化数据。

第 2 部分: RFE 特征融合与选择。将 LLM 提取的结构化指标与原有客户行为数据融合, 并利用 RFE 进行最优特征选择。

第 3 部分: XGBoost 最终预测。基于客服系统实际生产数据, 采用 XGBoost 对 RFE 选择的有效特征子集进行预测, 并通过基准模型对比说明该混合学习方法的有效性。

2 实证分析

2.1 实验环境

实验硬件配置为 12 核 CPU, 64 GB 内存, RTX 3060 显卡, 操作系统为 Win10, 编程环境为 Python3.12, 深度学习框架为 sklearn 1.6.1。

2.2 评估准则

因升级投诉客户与非升级客户比例差异较大, 存在数据分布不均衡的问题, 本文通过精准率 (Precision)、召回率 (Recall) 以及 F1-score (F1 值) 对不同模型的预测性能进行定量评估。其中, 精准率值越高说明预测为升级投诉的客户中, 实际升级的客户占比越高; 召回率越高说明实际升级投

诉的客户中, 被模型识别出来的比例越高; F1-score 是精确率与召回率的调和平均数, 适用于不平衡数据的综合评估。评估指标的具体描述如下:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{TP + FN} \quad (7)$$

其中, TP 表示预测为升级客户且实际升级的客户数; FP 表示预测为非升级客户但实际升级的客户数; FN 表示预测为升级客户但实际没有升级的客户数。

2.3 实证结果

为了验证所提出模型的有效性及其稳定性, 本文利用某省运营商客户实际生产中数据为例进行实证。该升级投诉数据集共 9 577 条记录, 字段包含呼入次数、通话时长、工单归属及处理满意度等 25 项特征, 以及目标变量客户是否升级投诉的标识。将数据按照 4:1 的比例随机划分为训练集和测试集, 其中训练集包含 7 625 条数据, 测试集含 1 951 条数据。

通过 5 折交叉验证在训练集内部严格进行递归特征消除, 每折独立使用随机森林剔除低重要性特征至保留 15 个, 并通过验证子集评估性能, 最终选出的特征在独立测试集上达到 F1 值 0.75, 表明该方法有效提升了模型泛化能力。最终从原始 25 个指标 (含 LLM 提取的 4 个指标) 中筛选出 15 个有效指标, 涉及人工接触、历史工单以及 LLM 从文本中挖掘的情感得分等, 选择特征及描述见表 3。

LLM 对文本数据以及工单内容处理后输出 4 个结构化字段, 与之前特征进行融合, 再利用 RFE 进行指标筛选, 最后将有效指标输入 XGBoost 进行预测。为了比较本文所提出的 RFE-LLM-XGBoost 框架的有效性, 选择朴素贝叶斯 (naive Bayes, NB)、随机森林 (random forest, RF)、轻量级梯度提升机算法 (light gradient boost-

ing machine, LightGBM) 3 个主流机器学习分类模型作为基准模型进行对比分析, 采用 Precision、Recall 和 F1 值 3 种评估指标对所有预测模型的性能进行评价。

表3 选择特征及描述

| 序号 | 特征名称 | 特征描述 |
|----|--------------------------|-----------|
| 1 | call_cnt | 呼入次数 |
| 2 | outbound_cnt | 外呼次数 |
| 3 | satisfaction_score | 人工场景满意度评分 |
| 4 | duration | 通话时长 |
| 5 | reminder_cnt | 催单次数 |
| 6 | order_cnt | 历史普通工单量 |
| 7 | unsatisfy_order_cnt | 工单不满意办结次数 |
| 8 | upcomplaint_cnt | 历史升级投诉量 |
| 9 | rule_complaint | 是否为业务规则投诉 |
| 10 | rule_complaint_cnt | 业务规则投诉量 |
| 11 | timeout_order | 超时在途工单 |
| 12 | order_satisfaction_score | 投诉场景满意度评分 |
| 13 | LLM_emotion | 情感得分 |
| 14 | LLM_demands | 关键诉求 |
| 15 | LLM_complaint | 升级投诉概率 |

不同模型的预测精度见表 4, 表 4 中为各个模型在升级投诉客户预警中的表现数据, 加粗字体为每列最优值, 结果表明:

(1) 在单一模型中, RFE-XGBoost 具有最优的精确率和召回率, 综合指标 F1 值也达到最大值, 这说明 XGBoost 算法更适用于升级投诉预警的场景。

表4 不同模型的预测精度

| 模型 | 精确率 | 召回率 | F1 值 |
|-----------------|--------------|--------------|--------------|
| RFE-NB | 87.7% | 51.2% | 0.654 |
| RFE-RF | 87.4% | 63.8% | 0.737 |
| RFE-Lightgbm | 88.3% | 70.0% | 0.781 |
| RFE-XGBoost | 88.4% | 72.9% | 0.799 |
| LLM-RFE-XGBoost | 90.1% | 75.2% | 0.820 |

(2) 通过引入 LLM 对用户来电文本进行情感等特征抽取, 再使用特征融合对模型进行增强, 可以有效提升小模型的预测性能。

(3) 相比 RFE-XGBoost, LLM-RFE-XGBoost 在精确率、召回率和 F1 值上分别提升了 1.9、3.2 与 2.6 个百分点, 引入 LLM 对用户来电文本进行情感特征和语义抽取后各个指标都有显著提高。验证了 LLM 在上下文语义理解方面的优势能有效弥补传统机器学习模型在上下文理解、情感极性捕获等任务上的固有局限, 两者结合可构建更具鲁棒性的集成预测方法。

3 应用效果

传统方法, 对于潜在升级投诉风险客户的识别主要依赖语音质检模型对风险关键字进行筛查, 随后由质检人员进行录音复听验证, 并对确认存在风险的客户实施回访补救。然而, 此方法面临识别量大、风险识别效率低等挑战, 难以快速精准地定位真正的升级投诉风险客户。升级投诉预警模型的引入有效解决了上述问题, 该模型基于客户呼入次数、工单量、工单处理满意度等多维度风险因子综合计算精准定位高风险客户, 输出命中风险投诉的明细信息, 并推送至投诉处理员开展服务回访补救, 可以辅助质量管理人员提前对升级投诉风险客户进行预警拦截。不仅减少了质检人员人工录音复听与核验的工作环节, 显著提升了潜在升级投诉风险客户的识别效率与准确率, 也大幅增强了升级投诉风险防控的预警拦截能力及处理质量。

目前, LLM-RFE-XGBoost 升级投诉预警模型已于某省客服中心上线应用, 2025 年 1 月—2025 年 5 月月均升级投诉量较 2024 年同期降低 6.7%, 效果显著。

4 结束语

本文提出的 LLM-RFE-XGBoost 混合学习框架, 通过本地化部署的 LLM 深度解析通话文本与工单内容, 提取相关特征; 融合原始结构化特征后, 采用递归特征消除 (RFE) 筛选高价值指



标, 最终输入 XGBoost 模型识别潜在升级投诉客户, 实现大小模型协同预警。基于某省运营商实际数据的实证分析表明: 该框架充分发挥了大模型在上下文理解与情感极性捕获等 NLP 任务上的强大能力, 以大模型作为高级特征提取器赋能轻量级 XGBoost 分类器可以显著提升模型预警精度, 更重要的是大小模型协同规避了大模型直接预测的高推理成本与结果不确定性问题。

综上所述, LLM-RFE-XGBoost 框架通过“大模型特征提取与小模型高效决策”的创新范式, 为非结构化数据挖掘提供了切实可行的技术路径, 其落地经验对于推动大模型在垂直领域的轻量化应用具有重要实践意义。

参考文献:

- [1] 景奕昕. 基于机器学习的电商平台客户投诉精准定位研究[J]. 数字通信世界, 2025(1): 22-24, 27.
Jing Y X. Research on precise positioning of customer complaints on e-commerce platforms based on machine learning[J]. Digital Communication World, 2025(1): 22-24, 27.
- [2] 王金政, 杨颖, 余本功. 基于多头协同注意力机制的客户投诉文本分类模型[J]. 数据分析与知识发现, 2023, 7(1): 128-137.
Wang J Z, Yang Y, Yu B G. Classifying customer complaints based on multi-head co-attention mechanism[J]. Data Analysis and Knowledge Discovery, 2023, 7(1): 128-137.
- [3] 祝好, 齐磊, 顾慧琼. 基于人工神经网络算法的运营商客户投诉智能分类问题研究[J]. 电信工程技术与标准化, 2021, 34(3): 31-35.
Zhu H, Qi L, Gu H Q. Research on intelligent classification of operator customer complaints based on artificial neural network algorithm[J]. Telecom Engineering Technics and Standardization, 2021, 34(3): 31-35.
- [4] 李青, 杨春霞, 蒋春辉, 等. 基于聚类分析和决策树的客户投诉预警体系[J]. 长江信息通信, 2022, 35(7): 180-183.
Li Q, Yang C X, Jiang C H, et al. Customer complaint early warning system based on cluster analysis and decision tree[J]. Changjiang Information & Communications, 2022, 35(7): 180-183.
- [5] 张艳莉. 基于 Logistic 回归算法的企业顾客抱怨管理机制的构建[J]. 科技通报, 2014, 30(5): 156-159.
Zhang Y L. Enterprise management mechanism for build customer complaints based on logistic regression algorithm[J]. Bulletin of Science and Technology, 2014, 30(5): 156-159.
- [6] 张蕊, 张丽红, 吴登群. 客户投诉处理精准分析与定位研究[J]. 数字通信世界, 2024(6): 58-60.
Zhang R, Zhang L H, Wu D Q. Research on accurate analysis and positioning of customer complaint handling[J]. Digital Communication World, 2024(6): 58-60.
- [7] 张爱华, 孙嘉鸿. 基于 TF-IDF 算法的运营商客户投诉原因研究[J]. 北京邮电大学学报(社会科学版), 2024, 26(2): 39-49.
Zhang A H, Sun J H. Reasons for customer complaints in operators based on TF-IDF algorithm[J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2024, 26(2): 39-49.
- [8] 郑杰明, 梁伟明, 陈震原, 等. 快速发现和定位投诉问题运用多元分析方法提升客户满意度[J]. 通信企业管理, 2023(11): 78-80.
Zheng J M, Liang W M, Chen Z Y, et al. Quickly find and locate complaints and improve customer satisfaction by multivariate analysis[J]. C-Enterprise Management, 2023(11): 78-80.
- [9] 周文杰, 严建峰, 杨璐. 基于深度学习的用户投诉预测模型研究[J]. 计算机应用研究, 2017, 34(5): 1428-1432.
Zhou W J, Yan J F, Yang L. Research on prediction model of complaint based on deep learning[J]. Application Research of Computers, 2017, 34(5): 1428-1432.
- [10] Brito J B G, Bucco G B, Heldt R, et al. A framework to improve churn prediction performance in retail banking[J]. Financial Innovation, 2024, 10(1): 17.
- [11] 马晓亮, 刘英, 高洁. 基于联合神经网络的投诉预测模型研究[J]. 电信科学, 2024(1): 48-58.
Ma X L, Liu Y, Gao J. Research on a complaint prediction model utilizing joint neural networks[J]. Telecommunications Science, 2024(1): 48-58.
- [12] 李龙戈, 郑铿城. 基于集成森林元学习网络的客户流失预测[J]. 电信科学, 2024, 40(10): 163-172.
Li L G, Zheng K C. Customer churn prediction based on the integration of meta-learning network of the forest[J]. Telecommunications Science, 2024, 40(10): 163-172.
- [13] 张爱华, 于涵, 陈超雨. 基于文本分析的运营商潜在升级投诉客户挖掘[J]. 北京邮电大学学报(社会科学版), 2022, 24(5): 32-38, 57.

Zhang A H, Yu H, Chen C Y. Mining of operators' potential customers with further complaints based on text analysis[J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2022, 24(5): 32-38, 57.

- [14] 穆晓玛. 主流客户投诉预测模型的对比与研究[J]. 智能计算机与应用, 2018, 8(5): 100-102, 105.

Mu X M. Comparison and research of several mainstream customer complaint prediction models[J]. Intelligent Computer and Applications, 2018, 8(5): 100-102, 105.

- [15] Wu D N, Zhou Z K, Su Z. Consumer negative opinions on stock returns: evidence from E-commerce reviews in China[J]. Financial Innovation, 2025, 11(1): 71.

- [16] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
Zhou Z H. Machine learning [M]. Beijing: Tsinghua University Press, 2016.

- [17] 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. New York: ACM Press, 2016.

[作者简介]



乔雅倩 (1996-), 女, 现就职于中移在线服务有限公司山东分公司, 主要研究方向为人工智能与数据挖掘。



王倩 (1987-), 女, 现就职于中移在线服务有限公司山东分公司, 主要研究方向为自然语言处理与数据挖掘。



刘芳 (1982-), 女, 现就职于中移在线服务有限公司山东分公司, 主要研究方向为智能客服解决方案、数据挖掘。



徐宇 (1974-), 男, 中移在线服务有限公司山东分公司副总经理、高级工程师, 主要研究方向为服务数智化、线上化运营。



江君 (1984-), 男, 就职于中移在线服务有限公司 AI 能力中心, 主要研究方向为智能客服、智能策略和智能体系统。