计算机科学
COMPUTER SCIENCE

基于XGBoost的电网物资供应商履约风险预测

李金霞, 卞华星, 温富国, 胡天牧, 秦诗涵, 吴涵, 马晖

引用本文

李金霞, 卞华星, 温富国, 胡天牧, 秦诗涵, 吴涵, 马晖. 基于XGBoost的电网物资供应商履约风险预测[J]. 计算机科学, 2024, 51(6A): 230400115-9.

LI Jinxia, BIAN Huaxing, WEN Fuguo, HU Tianmu, QIN Shihan, WU Han, MA Hui. Performance Risk Prediction of Power Grid Material Suppliers Based on XGBoost [J]. Computer Science, 2024, 51(6A): 230400115-9.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230400115

基于 XGBoost 的电网物资供应商履约风险预测

李金霞[1]　卞华星[1]　温富国[1]　胡天牧[2]　秦诗涵[3]　吴　涵[3]　马　晖[3]
1 国网江苏省电力有限公司物资分公司　南京 210036
2 江苏电力信息技术有限公司　南京 210000
3 中国科学院信息工程研究所　北京 100085
(798885953@qq.com)

摘　要　电网物资供应商履约质量是电网安全稳定运行的基础,供应商履约涉及环节众多且风险因素复杂,使得当前对其研究较为匮乏且大多停留在理论业务分析层面。针对这一问题,提出基于 XGBoost(Extreme Gradient Boosting)的供应商履约风险预测模型,充分考虑业务全流程中的各种风险因素,综合内部供应链运行、知识图谱数据以及外部天眼查、疫情等数据,基于特征工程构造了 191 个风险特征进行初始训练,在模型优化后对筛选出的 49 个特征再次训练,兼顾实际业务中的预测准确性和特征可解释性要求,采用 SHAP(SHapley Additive exPlanations)值方法进行模型解释。实验结果表明,对比其他 3 种主流机器学习算法,所提模型准确率、精确率、KS 值分别高达 93.05%,94.45%,45.38%,进而验证了 XGBoost 算法在履约风险预测中的可行性和优越性。该模型可应用到电网物资供应链中,进一步指导业务应用。

关键词:XGBoost;特征工程;供应商履约;风险预测;电网
中图分类号　TP309

Performance Risk Prediction of Power Grid Material Suppliers Based on XGBoost

LI Jinxia[1],BIAN Huaxing[1],WEN Fuguo[1],HU Tianmu[2],QIN Shihan[3],WU Han[3] and MA Hui[3]
1 State Grid Jiangsu Electric Power Co.,Ltd. Materials Branch,Nanjing 210036,China
2 Jiangsu Electric Power Information Technology Co.,Ltd,Nanjing 210000,China
3 Institute of Information Engineering,Chinese Academy of Sciences,Beijing 100085,China

Abstract　The performance quality of power grid material suppliers is the basis for the safe and stable operation of power grid, which involves many links and complex risk factors, causing the current research on it is relatively scarce and stays at the level of theoretical analysis. In order to solve this problem, a supplier performance risk prediction model based on XGBoost is proposed, which fully considers various risk factors during the whole process, integrates internal supply chain operation, knowledge map data, external eye inspection, epidemic situation and other data, constructs 191 risk features based on feature engineering for initial training, and retrains 49 selected features after model optimization, taking into account the requirements of prediction accuracy and feature interpretability in actual business, and uses SHAP method to explain the model. Experimental results show that, compared with other three mainstream machine learning algorithms, the accuracy rate, precision rate and KS value are as high as 93.05%, 94.45% and 45.38%, which further verifies the feasibility and superiority of XGBoost model in the performance risk prediction. The prediction model can be applied to the power grid supply chain business to further guide the practical application.

Keywords　XGBoost, Feature engineering, Supplier performance, Risk prediction, Power grid

1　引言

近年来,电网企业在现代工业经济发展中占据的地位越来越重要[1]。电网物资作为电网建设与维护的基础,其能否如期供应、质量好坏以及成本高低,极大地影响着整个电网系统的稳定、安全与高效发展[2]。电网物资供应商作为物资供应与质量保障的责任主体,能否按照合同及时履约,是影响电力物资能否按时供应以及质量好坏的主要风险来源,进而影响着电网安全运营的稳定性和可靠性。然而,电力物资种类多样且金额较大、供应涉及环节众多、合同履约时间长[3],加之受新冠疫情等外部因素影响,供应商违约行为在数量上呈现上升趋势。因此,面对电力物资供应履约的多重信用风险,研究影响供应商履约行为的各种风险因素并对其潜在的履约风险精准防控具有十分重要的意义。

供应商履约风险是指按照合同约定、缔约主体和项目物资合同履约的实际结果之间存在义务与职责预期目标偏差的可能性[4]。在合同签订后,供应商履约风险贯穿在需求计划、采购中标、合同签订、供应履约、出厂试验、物资发运、到货交接等全流程中。同时,造成供应商履约风险的因素众多,例如,供应商在合同履约过程中存在项目管理不当、公司财务信

基金项目:国网江苏省电力有限公司科技项目(J2022071)
This work was supported by the Science and Technology Project of State Grid Jiangsu Electric Power Co.,Ltd(J2022071).
通信作者:卞华星(nuaa_bhx@sina.com)

用风险、生产经营风险、交易信用风险等多种内部因素，同时原材料供应、疫情、天气等外部因素的影响，极大可能导致违约情况的发生。因此，供应商履约风险预测的目标是：综合影响供应商履约的多维度因素，考虑物资供应全流程的动态变化，精准评估供应商在不同环节的履约风险。

目前，电网物资供应商履约风险预测方法主要以业务经验分析结合统计学方法为主，其研究结果往往缺乏充足数据的支撑和科学方法的验证。Li 等[5]根据业务数据分析及有关管理规定，分析影响履约风险发生频率和发生等级的相关因素以及相应的关联度，设计了一种基于模糊层次分析法的供应商履约风险预警模型，但该模型仅为理论研究且过于依赖业务分析。Qu 等[6]提出一种多数据源融合的电力物资供应商评价知识图谱构建流程，实现供应商评价与监造策略的智能匹配，但其中供应商绩效指标权重使用专家评分法，该方法过于主观且无严谨的实验验证。Chen 等[7]通过指标选择、数据清洗、特征工程、体系构建等过程，构建了基于 XGBoost 算法的电力供应商及客户价值体系，供应商履约风险仅作为其供应商评价的一部分，并未考虑全流程的供应商履约风险预测工作。Shi 等[8]分析了新冠疫情出现后电力物资供应商履约存在的问题及相应解决策略，然而此研究仅为理论分析，缺乏数据支撑和实验验证。

基于以上工作可知，目前电网供应商履约风险预测工作相对匮乏，且多集中在供应商评估方面，所采用方法多为理论分析、统计学方法、专家打分法等，未能充分考虑业务全流程中各种风险特征并挖掘数据深层的规律，尚无完备的数据集支撑及验证。同时，其他领域的模型预测工作，例如微博流行度预测[9]、短期电力负荷预测[10-11]、小微企业违约特征评估[12]、足球运动员身价分析[13]等，充分验证了机器学习算法在预测模型中的可行性和优越性。然而，与传统机器学习算法不同的是，电网物资供应商履约风险预测模型缺乏完备的数据集，其工作难点是需要根据业务特点，分析供应商履约涉及各环节的风险特征，并对数据进行分析、清洗以及特征工程等工作，以获得数据集，进而开展模型训练及实验验证。

本文选取机器学习模型中兼具准确性与计算效率的 XGBoost 模型，提出了基于 XGBoost 的电网供应商履约风险预测模型，结合实际业务分析，首先通过特征工程构造了 191 个风险特征进行初始训练，通过模型优化后，对筛选出的 49 个特征再次训练，并采用 SHAP 值方法进行模型解释，通过将模型输出结果转化为具体概率，可进一步指导实际应用，辅助相关业务人员决策。本文贡献及创新点主要包括：

(1)针对当前电网供应商履约风险预测工作匮乏的现状，对业务原始数据深入挖掘，并建立特征工程，进一步形成完备的电网供应商履约风险数据集。

(2)建立了基于 XGBoost 的电网供应商履约风险预测模型，通过数据清洗、特征工程、模型训练、优化、评估、解释、应用等 7 个过程验证了模型的有效性和准确性。

(3)通过 SHAP 值方法增强了本文模型结果的可解释性，精准识别影响供应商履约的核心因素，提高了模型的实用性，进而对电网业务应用提出有效的指导意见。

2 基于 XGBoost 的电网供应商履约风险预测模型

本文基于 XGBoost 的电网供应商履约风险预测模型架构如图 1 所示，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、模型应用等 7 个流程。下面结合电网物资供应商履约的具体业务流程介绍上述流程所采用的方法和理论依据。

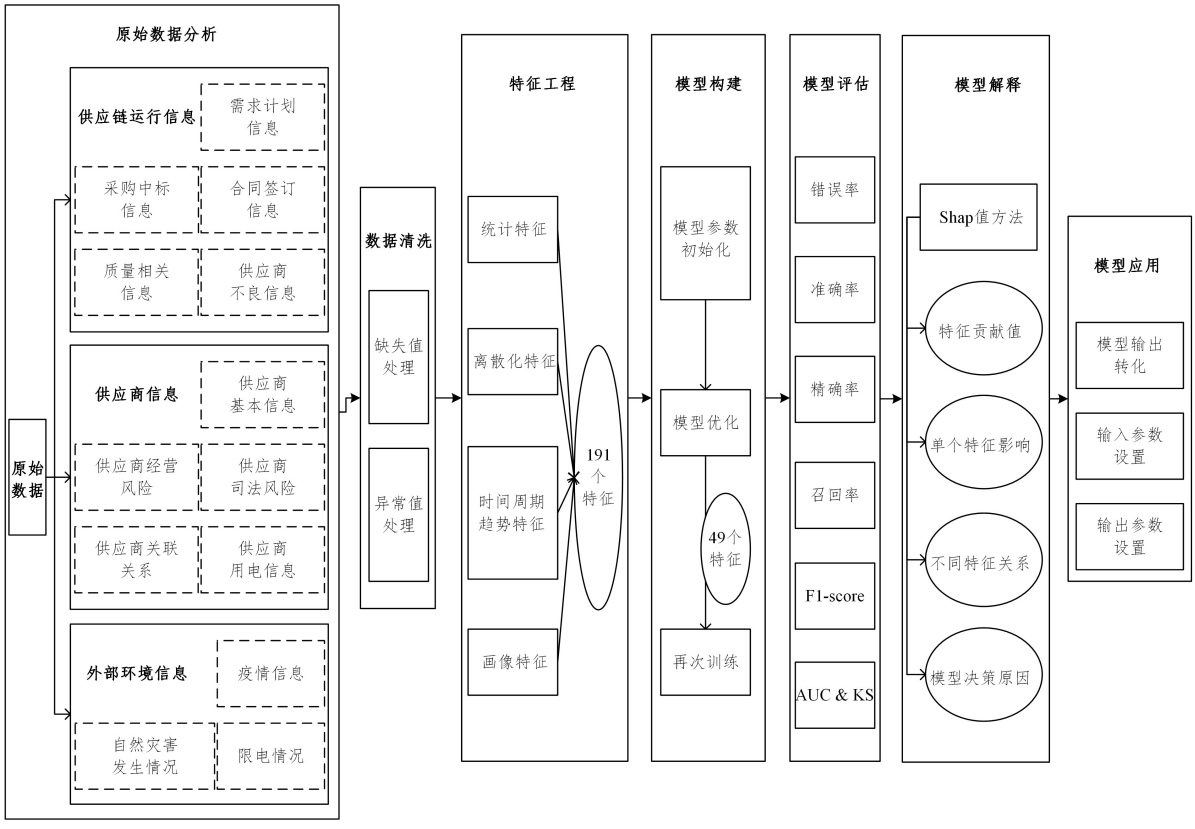

图 1 基于 XGBoost 的电网供应商履约风险预测模型架构

Fig. 1 Framework of contract performance risk prediction model for power grid suppliers based on XGBoost

2.1 原始数据分析

本文原始数据来源于江苏电力物资公司供应合同履约数据库、知识图谱数据库、天眼查平台以及中国卫健委网站，包括物资履约信息表、物资未完成供应数据表、供应商生产质量信息预警表等 8 个表。基于实际业务分析，将以上原始数据划分为 3 类：供应链运行信息、供应商信息和外部环境信息。

供应链运行信息中的数据项包括：物资唯一码、物料编码、物料组、供应商编码、是否价格联动、分标编号、分包编号、中标单价、中标总价、采购任务接受时间、合同交货期、实际交货期、合同签订日期、计划提报日期、是否抽检、抽检结果是否合格、原材料类别、原材料价格等。

供应商信息包含数据项：法律诉讼、经营异常、欠税公告、清算信息、失信公司、税收违法、司法协助、土地抵押面积、严重违法等。

外部环境信息包含数据项：发生疫情城市、累计确诊、是否发生自然灾害、是否限电等。

本文预测模型的任务是风险评估与预测，首先应选取合适的分析粒度。通过对以上原始数据进行分析，我们选择其中最小的粒度“物料”作为分析对象。因为在履约数据库中是以物料为单位记录违约数据，而且每个物料都对应着一个物料唯一码，该码既能够链接到合同等供应链运行信息，又能够链接到供应商，有助于对供应商履约的各个环节实现风险预测及管控。

2.2 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。

2.2.1 缺失值处理

由于业务原因，原始数据中部分字段存在数据缺失的问题，需要进行缺失值处理的相关字段、原因以及处理方法如表 1 所列。

表 1 缺失值字段、原因及处理方法

Table 1 Missing and abnormal value fields, reasons and processing methods

字段	原因	处理方法
是否价格联动、分标编号、分包编号	非协议履约合同，缺少相应字段	默认填充为“N”
实际交货期	截止至统计日期仍未完成交货	自动判断为 missing 条件
司法协助、司法协助股权数额	供应商公司违法概率较低	自动判断为 missing 条件
交货时累计确诊、订单接受时累计确诊等	履约时无疫情	默认填充为“0”

2.2.2 异常值处理

原始数据在采集、加工、传输等流程中产生的异常数据，容易造成下游任务的数据倾斜和报错，影响模型鲁棒性，因此需要进行异常值处理。例如，在物资供应生命周期中，部分日期数据存在早于上一节点或晚于下一节点等异常问题，将其标注为异常值，并取平均日期。

2.3 特征工程

供应链业务本身涉及环节众多，影响供应商履约因素复杂，各类特征重要性各有差别，因此进行特征的提取与构造十分必要。本文采用以下特征工程方法，构造了 191 个风险特征进行模型初次训练。

2.3.1 统计特征

由于履约数据大多为明细数据，例如价格、金额、数量，通过统计量的方式，可以快速地汇总并刻画出供应合同数据的分布规律。我们对每个数值开展了包括总和、最大值、平均值、比例、排名、频率、近 3 个月、近 6 个月、近 12 个月等多角度的聚合计算。

2.3.2 离散化特征

由于供应商所在城市、公司类型等字符型数据无法直接进行模型训练，因此需对其进行离散化，即将连续值转化为离散值。离散化之后的特征稳定性增强的同时，能够提高模型的泛化能力。

离散化主要包括两种编码方法：标签编码（Label Encoding）和独热编码（One-Hot Encoding）。对比两种离散化方式，独热编码会导致特征矩阵的稀疏性，并且会损失一些数据本身的信息；此外，在对数值大小变化不敏感的树形模型中，标签编码在完成独热编码功能的同时可以进行排序和变量合并，一般具有更好的效果。因此，我们选择对字符型数据进行标签编码。

2.3.3 时间周期趋势特征

在电网供应链业务中，供应商供应累计时间较长，且在时间上的分布并不稳定，因此我们加入考虑时间距离的特征，包括年均违约金额、价格方差等，这类时间周期趋势特征能够天然地刻画出供应商中长期的供应履约风险。

2.3.4 画像特征

考虑到供应商所在公司的经营情况对履约概率具有一定影响（例如公司破产无法完成履约），从外部数据源引入供应商的画像特征，包括经营异常、司法风险、注册资本等。此类画像特征的引入既可以增强预测模型的可解释性，也可以在满足合规要求后对外输出。

原始数据中某些特征单独来看可能和风险关联性并不强，但通过不同时间段划分、多种统计方式处理后，却会与预测目标产生紧密关联。基于以上特征工程方法，构造出 191 个风险特征：样本计划提报周期、二级供应商关联关系风险总数、合同执行周期、物料单价、历史物料单价平均值、月均供应数量、年均供应数量、近 12 个月/历史年均订单数、近 3 个月/近 6 个月/近 12 个月/历史供应金额方差、近 3 个月/近 6 个月/近 12 个月/历史违约金额最大值、近 3 个月/近 6 个月/近 12 个月/历史违约次数、供应商历史抽检不合格次数等。此处由于文章篇幅有限，不进行一一列举，仅在表 2 中展示影响供应商履约贡献度较高的前 49 个特征。其中，特征贡献度的评估基于 SHAP 值方法，以特征的权重总和占比＞90％是为阈值作为筛选标准（此处设定阈值为 90％依据主观决策和实验验证，经过多次实验调整后我们发现，加入特征权重总和占比不高于 10％的若干个特征，模型将会过拟合，使得其 KS 值下降）。

表 2 贡献度较高的前 49 个特征列表

Table 2 Top 49 features with high contribution

序号	特征类别	特征名称	特征说明
1	供应链运行信息	计划提报周期	计划提报周期=计划交货日期-计划提报日期
2		所在年份计划提报周期平均值	计划提报日期所在年份供应记录中具有相同物料编码的物料的计划提报周期的平均值=物料编码的计划提报周期总和/订单数
3		历史计划提报周期平均值	往年供应记录中具有相同物料编码的物料的计划提报周期的平均值=物料编码的计划提报周期总和/订单数
4		计划提报周期孤立值	(计划提报周期-历史计划提报周期均值)/历史计划提报周期均值
5		供应商供货聚集性	供应商当前(该需求计划提报日期)正在执行的合同订单数
6		合同执行周期	合同执行周期=合同指定交货期-合同签订日期
7		物料单价	合同中的物料单价
8		历史物料单价平均值	合同签订日期前的历史供应记录中具有相同物料编码的物料单价平均值=单价总和/订单数
9		历史物料单价方差	合同签订日期前的历史供应记录中具有相同物料编码的物料单价的方差=(单价-单价均值)平方之和/合同数
10		物料总价	合同中的物料总价
11		历史物料总价平均值	合同签订日期前历史供应记录中具有相同物料编码的总价平均值=总价总和/订单数
12		历史物料总价方差	合同签订日期前历史供应记录中具有相同物料编码的总价的方差=(总价-总价均值)平方之和/合同数
13		历史物料总价排名	物料总价在合同签订日期前历史供应记录中具有相同物料编码中的排名(降序)
14		合同的物料数量	合同的物料数量
15		历史物料数量最大值	合同签订日期前历史供应记录中相同物料编码的物料数量最大值
16		历史物料数量平均值	合同签订日期前历史供应记录中具有相同物料编码的物料数量的平均值=数量总和/订单数
17		历史物料数量方差	合同签订日期前历史供应记录中具有相同物料编码的物料数量的方差=(数量-物料编码数量均值)平方之和/合同数
18		分包最高单价	所属分标下所有分包的最高单价
19		分包平均单价	所属分标下所有分包的平均单价=分包内单价总和/分包订单数
20		分包单价方差	所属分标下所有分包的单价方差=(分包单价-分包单价均值)平方之和/合同数
21		分包单价排名	在所属分标下的所有分包中的单价排名(降序)
22		最早供应距今时长	供应商供应历史记录中最早一次供应记录至今(2022.9)的天数
23		历史违约金额最大值	供应商违约历史中供应合同金额的最大值
24		历史违约金额平均值	供应商违约历史中供应合同金额的平均值=合同金额总和/合同数
25		历史违约金额总和	供应商违约历史中供应合同金额的总额
26		历史违约金额占比	供应商历史违约合同金额的总额/历史供应合同金额总额
27		历史违约次数	供应商近历史违约次数
28		历史违约率	供应商历史违约次数/历史供应次数
29		历史履约金额占比	供应商历史履约合同金额的总额/历史供应合同金额总额
30		历史履约次数	供应商近历史履约次数
31		历史履约率	供应商历史履约次数/历史供应次数
32		原材料类别	原材料类别
33		供应期间(供货单创建至预计交货期)原材料单价平均值	供应期间原材料单价的平均值=供应期间原材料单价总和/月份数
34		供应期间原材料单价方差	在供应期间原材料单价的方差值=(原材料单价-原材料单价均值)平方之和/合同数
35		是否抽检	物料是否被抽检
36		抽检结果是否合格	物料抽检结果是否合格
37		供应商历史抽检不合格次数	供应商抽检结果为不合格的次数
38		供应商历史不合格率	供应商的历史抽检不合格次数/历史抽检次数
39		历史处罚期限总和	一级供应商的历史处罚期限的总和
40		历史处罚次数	一级供应商的历史触发次数
41	供应商信息	注册资本	一级供应商的注册资本金额
42		欠税余额	一级供应商的欠税余额
43		经营异常次数	一级供应商的经营异常次数
44		土地抵押面积	一级供应商以及工商的土地抵押面积
45		失信人数量	一级供应商的失信人数量
46		法律诉讼次数	一级供应商的法律诉讼次数
47		司法协助次数	一级供应商的司法协助次数
48	外部环境信息	订单接受时累计确诊	供应商接受该订单时供应生产地的累计确诊人数
49		交货时累计确诊	供应商交货时供应目的地的累计确诊人数

2.4 模型构建

本文模型构建包括模型参数调整、模型训练以及模型优化，在模型优化前后分别进行了两次模型训练。

2.4.1 模型参数初始化

为有效防止模型过拟合、数据不平衡以及特征征浓度不平衡问题，需要进行模型参数调整，为 XGBoost 模型选出一组最优超参数组合。

目前主流参数调整方法包括网格搜索、随机搜索以及贝叶斯优化。其中，贝叶斯优化是 XGBoost 监督模型学习中调整参数和提高模型泛化性能的有效方法，可以避免网格搜索

带来的维度灾难；同时，贝叶斯优化在超参数调整中相较于随机搜索具有更高的效率、自适应性和有效性，更能够准确地找到最优的超参数组合。因此，本文选择贝叶斯优化来调整模型参数，以获得最优性能。

为进一步实现自动化超参数调参，节省参数选择时间和资源，本文选择采用 Hyperopt。Hyperopt 是一个基于 Python 实现的贝叶斯优化调参工具，在超参数搜索过程中可以高效地利用之前的结果并自适应地探索搜寻空间，从而加速调参过程。其优化过程包含：(1)初始阶段：随机采样生成一组超参数来训练模型，并记录其性能指标(如准确率、F1-score 等)；(2)建立模型：利用这些初始数据建立高斯过程回归(Gaussian Processes，GP)模型，该模型会将每个超参数组合作为输入，并输出由历史记录决定的预测性能值及其方差，以此来估计函数的概率分布情况；(3)选择新的采样点：通过利用贝叶斯公式，结合当前已有的数据和 GP 模型的预测结果，计算出目标函数可能达到最大值的超参数组合，并根据目标函数在该超参数组合处的期望方差大小确定下一个要采样的点；(4)评估：使用所选超参数组合来训练模型，并记录其性能指标；(5)更新 GP 模型：将新获得的数据点加入历史记录中，重新训练 GP 模型，更新超参数空间的概率分布估计；(6)迭代上述步骤，直至达到预先设定的迭代次数或者超参数调整目标函数的最优解符合要求。Hyperopt 已经在实际问题中展现了很好的效果，并成为了许多算法工具箱中调参不可缺少的部分。

2.4.2 模型训练

得到超参数值最优组合后，将其代入 XGBoost 模型中训练。极端梯度提升(Extreme Gradient Boosting，XGBoost)是一种基于梯度提升树和集成学习的模型[14]，是树模型中兼具准确性与计算效率的主要模型之一，因此本文选择 XGBoost 作为风险预测和实验验证过程所使用的模型。

本文算法架构中，输入特征表示为：$I=\{(f_i,p)\mid i=1,2,\cdots,n,f_i\in R^d,p\in 0,1\}$，其中 f_i 表示履约数据的全部特征，p 表示该记录是否违约，n 表示训练样本的总数，d 表示特征的维度，R^d 表示 d 维实数集。模型训练目的是输出预测的合同履约概率 $\hat{p}_l$，其定义如式(1)所示：

$$\hat{p}_l=\phi(f_i)=\sum_{k=1}^{K}g_k(f_i),\hat{p}_i\in\Gamma \tag{1}$$

其中，k 表示激活函数，K 表示激活函数的总数，g 表示权重为 ω 的分类决策树，Γ 表示 XGBoost 分类树的映射空间，映射关系如式(2)所示：

$$\Gamma=\{g(x)=\omega f(x)(f:R^d\to L,\omega\in R^L) \tag{2}$$

其中，ω 表示权重，每个独立树 f 包含 L 个子树，R^L 表示 L 维实数集。为了优化学习效果，确定 XGBoost 的初始目标函数如式(3)所示：

$$obj=\sum_{i=1}^{n}l(y_i,\hat{y}_i^{(t)})+\sum_{i=1}^{t}\omega(f_i) \tag{3}$$

其中，前一项是训练损失，n 表示训练样本的总数，$\hat{y}_i$ 是预测值，y_i 是真实值，l 是损失函数；后一项是用于防止过拟合的正则项。由于一次训练完所有树并不容易，因此逐步添加新树，通过 $\hat{y}_i^{(t)}$ 来关注步骤 t 的预测值：$\hat{y}_i^{(t)}=\sum_{k=1}^{t}f_k(x_i)=\hat{y}_i^{(t-1)}+f_t(x_i)$。本研究使用 MSE 作为损失函数，并将其泰勒展开到二阶，如式(4)所示：

$$obj^{(t)}=\sum_{i=1}^{T}[g_iw_i+\frac{1}{2}(h_i+\lambda)w_i^2]+\gamma T \tag{4}$$

其中，γ 表示剪枝参数，λ 表示正则化的超参数，T 表示叶子总数，g_i 和 h_i 分别被定义为：

$$g_i=\partial_{\hat{y}_i^{(t-1)}}l(y_i,\hat{y}_i^{(t-1)}) \tag{5}$$

$$h_i=\partial^2_{\hat{y}_i^{(t-1)}}l(y_i,\hat{y}_i^{(t-1)}) \tag{6}$$

g_i，h_i 分别为第 i 个处损失函数的一阶导数与二阶导数。

2.4.3 模型优化

为了进一步提高本文预测模型的性能，我们从参数优化和数据优化两方面进行模型优化。

(1)参数优化

优化内容：调整 gamma 指定叶节点进行分支所需的损失减少的最小值；由于模型特征数较多，因此调整 alpha L1 正则化权重项和 lambda L2 正则化权重项，使模型更加保守，防止过拟合；调整 max_depth 指定树的最大深度，防止过拟合；调整 min_child_weight 指定子节点中最小的权重和；指定 subsample 指定子采样用于训练弱学习器，防止过拟合；调整 colsample_bytree 在构建弱学习器时对特征随机采样的比例。

优化方法：与模型参数调整相同，本文采用贝叶斯优化结合 Hyperopt 来实现自动化超参数调整。

(2)数据优化

优化内容：由于大多数供应商不存在司法风险，因此对应数据集中的数据存在大量缺失值，导致特征浓度不平衡。

优化方法：本文通过对大类数据进行欠采样，放弃部分大类数据来实现优化。

2.5 模型评估

本文使用受试者工作特征曲线(Receiver Operating Characteristic Curves)下面积 AUC(Area Under Curve)、KS(Kolmogorov-Smirnov)检验值及其他 5 个指标评估模型性能。

2.5.1 AUC 值

AUC 为受试者操作特征曲线下面积，该曲线图是反映敏感性与特异性之间关系的曲线。横坐标 X 轴为 1-特异性，也称假阳性率(误报率)，X 轴取值越接近零准确率越高；纵坐标 Y 轴称为敏感度，也称真阳性率(敏感度)，Y 轴取值越大准确率越好。

2.5.2 KS 检验

KS 检验是一种基于累计分布函数的非参数检验(Empirical Cumulative Distribution Function，ECDF)，用以检验两个经验分布是否不同或一个经验分布与另一个理想分布是否有显著性差异。统计量 KS 值即是 KS 检验的结果。对于集合 X 及其累计分布函数 $F_n=\frac{1}{n}\sum_{i=1}^{n}I_{[-\infty,x]}(X_i)$ 和假设的理论分布 $F(X)$，统计量定义为：$D_n=\sup_x|F_n(x)-F(x)|$。

首先按自定义距离或频率对变量进行分箱，之后计算每个分箱区间内累计 A 类数占总 A 类数的比率，以及 B 类同样的比率，之后计算每个分箱中这两个比率差的绝对值，最后取这些绝对值的最大值，即为最终的 KS 值，即

$KS=\max\{|cum(A)-cum(B)|\}$。KS指标越大，模型的风险区分能力越强。

2.5.3 其他常用指标

除此之外，我们还引入了以下5个常用指标：错误率(Errorr)、准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F_1评分(F_1-score)。其计算方法分别如下：

$$Error=\frac{FP+FN}{TP+FN+TN+FP}=\frac{FP+FN}{P+N} \tag{7}$$

$$Accuracy=\frac{TP+TN}{TP+FN+TN+FP}=\frac{TP+TN}{P+N} \tag{8}$$

$$Precision=\frac{TP}{TP+FP} \tag{9}$$

$$Recall=\frac{TP}{TP+FN}=\frac{TP}{P} \tag{10}$$

$$F_1=2\times\frac{precision\times recall}{precision+recall} \tag{11}$$

其中，TP，FP，TN，FN分别表示真阳性、假阳性、真阴性、假阴性。

2.6 模型解释

本文选择基于事后解释框架的SHAP(SHapley Additive exPlanations)值方法[15]进行模型解释，并选取TreeSHAP作为实现方法。SHAP值方法能够很好地解释本文构建的XGBoost供应商履约风险预测模型中特征的贡献程度，明晰每个变量与目标之间的正负向关系。对于每个预测，模型都会产生一个预测值，SHAP值就是该样本中每个特征对应的数值。假设第i个样本为x_i，第i个样本的第j个特征为x_{ij}，模型对第i个样本的预测值为y_i，模型的基线(通常是所有的目标变量的均值)为y_{base}，那么SHAP值满足：

$$y_i=y_{base}+f(x_{i,1})+f(x_{i,2})+\cdots+f(x_{i,k}) \tag{12}$$

其中，$f(x_{i,j})$为$x_{i,j}$的SHAP值。可以看出，当SHAP值为正时，该特征对目标的贡献度为正向；反之则为负向。

2.7 模型应用

为方便将本文提出的供应商风险预测模型应用到实际业务中，本文考虑将模型输出结果“0/1”转化为更直观的履约风险概率。因此，我们选择使用XGBClassifier中的predict_probe()将输出结果转化为[0,1]区间内的概率，再通过分数映射将概率转化为[0,1000]区间内的分数，方便设定阈值来拦截。

3 实验结果分析

实验硬件平台处理器为Intel i7-78720HQ，内存为16 GB，固态硬盘容量为500 GB，GPU显卡为Radeon 560。软件平台为DataSpell，使用基于XGBoost的机器学习库。

本文共进行两次模型训练。第一次模型训练中，我们使用特征工程输出的191个特征来训练，通过模型参数优化结合SHAP值贡献程度筛选出优化后的49个重要特征。经过初步筛选得到的49个特征均具有一定的区分度，若进行再特征降维可能会损失一些信息熵，不利于整体性能，因此利用这49个特征进行第二次模型训练。模型性能评估结果表明，优化后的模型各项指标都有显著提升。使用TreeSHAP方法对实验结果进行多维度解释，并选取多种模型开展对比实验，验证了XGBoost模型的优越性。

3.1 实验数据集

本文实验原始数据共包含供应合同履约数据66 621条，其中履约记录59 999条，占比90.06%，违约记录6 622条，占比9.93%；新冠病毒疫情确诊数据时间跨度为2019年12月到2022年10月，包含确诊数量的履约记录37 129条，占比55.73%。通过特征工程，输出191个影响供应商履约的风险特征，进而得到本文实验数据集。使用sklearn包的train_test_split函数进行数据集划分，使用默认的test_size参数=0.25，即训练集占比为75%，测试集占比为25%。

对于数据集中量纲不一致的数据，通过sql中的字符串函数配合正则表达式和sklearn库中的函数进行处理。例如金额相关数据中，注册资本存在“xx元”“xx美元”等不一致的形式，使用字符串函数和正则函数抽取后进行归一化处理。

3.2 超参数优化结果

按照2.4节所述方法对模型进行参数调整和优化，优化前后的XGBoost超参数的最优组合如表3所列。将两组超参数最优组合分别代入XGBoost模型，使用训练集完成两次训练。

表3 优化前后的XGBoost超参数的最优组合

Table 3 Optimal hyperparameters of XGBoost before and after optimization

超参数名称	优化前取值	优化后取值
colsample_bytree	1	1
eta	0.7	0.7
gamma	2.722	4.905
max_depth	8	5
min_child_weight	1	1
n_estimators	802	350
reg_alpha	56	52
reg_lambda	0.89	0.66
scale_pos_weight	2.3	1.9
subsample	0.9	0.6

3.3 模型性能评估

在测试集上对模型优化前后进行性能评估，模型优化后各项指标都有明显提升，结果见表4。优化前后的模型都具备良好的供应商履约风险预测能力，进一步验证了XGBoost模型应用于供应商履约风险预测的可行性和有效性。但从AUC值和KS值来看，初始训练模型存在有效性低($AUC<80\%$)、区分度不足($KS<10\%$)问题，通过模型优化，这两项指标分别提升65.38%和501.86%，有效缓解了特征权重失衡的问题。优化后模型预测结果的AUC值为0.86，验证了优化后的预测模型更具有实际应用价值。

表4 优化前后的性能对比

Table 4 Performance comparison before and after optimization

(%)

指标	优化前数值	优化后数值	提升幅度
Error	0.13	0.07	46.15
Accuracy	86.76	93.05	7.25
Precision	90.25	94.45	4.65
Recall	95.64	99.01	3.52
F_1-*score*	92.88	96.22	3.60
AUC	52	86	65.38
KS	7.54	45.38	501.86

3.4 模型实验结果解释

本文使用 TreeSHAP（TreeSHAP 是 SAHP 值方法的一种高效实现方法）对模型实验结果进行解释，分别对各项特征对模型预测结果的贡献、单个特征对模型预测结果的影响、不同特征之间的关系以及模型决策背后的原因进行解释和分析。

3.4.1 各项特征对模型预测结果的贡献解释

基于 SHAP 值方法计算了本模型中的每一项特征对预测结果的贡献值，从而判断哪些特征是影响供应商履约的重要特征。贡献较高的前十个特征的 SHAP 值排序如表 5 所列。

表 5 重要性排序前十特征列表

Table 5 Top 10 features with high contribution

特征排序	特征名称	SHAP 值
1	历史违约率	0.71
2	合同执行周期	0.50
3	交货时累计确诊数	0.33
4	历史履约率	0.27
5	历史计划提报周期均值	0.21
6	计划提报周期	0.18
7	物料总价	0.15
8	物料单价方差	0.15
9	历史违约金额均值	0.11
10	分包单价排名	0.09

上述 10 个特征的特征全局解释如图 2 所示。综合特征取值、SHAP 值以及多特征呈现，充分解释了特征对预测目标的贡献度和贡献方向。其中，纵坐标为特征名称，横坐标为 SHAP 值，每个点都是一个样本在该特征下的 SHAP 值。SHAP 值大于 0 表示正贡献，即该特征取值越大，最终违约概率越小；反之亦然。特征在点上的取值大小采用不同颜色表示，红色点代表特征在该样本中取值较高，蓝色点代表取值较低。如图 2 所示，历史违约率对供应商履约风险影响最大，正贡献极小而负贡献极大，所以供应商的历史记录中的违约率越大，违约的概率越高；合同执行周期是第二重要的特征，合同执行周期越短，发生违约的概率越低。上述实验结果与业务经验结果相一致。

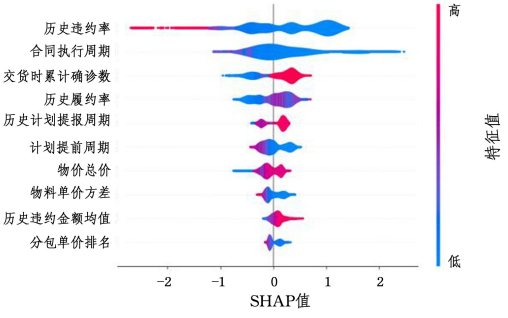

图 2 特征全局解释图

Fig. 2 Global interpretation of features

3.4.2 单个特征对模型预测结果的影响分析

为研究单个特征对模型预测结果的影响，对每个特征分别绘制其 SHAP 值散点图，直观地展示单个特征的取值变化对模型预测结果的影响趋势。

以历史违约率这一重要特征为例，其 SHAP 值散点图如图 3 所示，图片中蓝色部分表示历史违约率不同取值对模型预测结果的不同贡献值，图片底部的浅灰色区域是数据值分布的直方图。

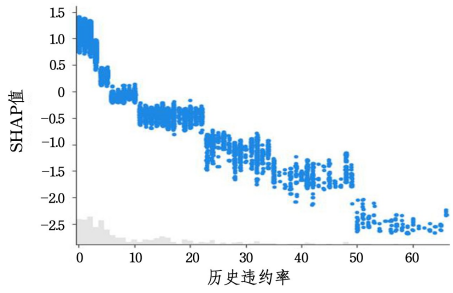

图 3 历史违约率的 SHAP 值散点图

Fig. 3 Scatter chart of SHAP value of historical default rate

由图可知，当供应商的历史违约率高于约 6% 后，其对模型预测结果的贡献由正转负，供应商历史记录中违约的概率越高，供应商能够履约的概率越低。这一结论与已有经验相一致，从供应商历史违约情况来看，其违约原因一般涉及原材料价格波动、疫情增长等外部原因，也包含企业实施管理不当、经营不善等内部问题。因此，本文建议电网供应链业务人员在实施电网物资建设项目时重点考察供应商历史记录中的违约概率，及时防范履约风险。

3.4.3 不同特征之间的关系分析

为进一步探究供应商履约风险预测模型中不同特征之间的关系，可选择多个特征综合绘制其 SHAP 值散点图，帮助电网供应链业务人员更好地了解各特征之间的交互效应。

以疫情增长与供应商历史违约率的交互关系为例，如图 4 所示，我们选择疫情增长作为特征来确定历史违约率从 0 增加到 60 时的影响。红色的点代表历史违约率的较高值，蓝色的点代表较低值。从图 4 中可知，在疫情增长的作用下，供应商历史违约率产生了一定程度的波动，说明这两个特征间存在非线性的交互作用，即疫情增长严重时，供应商历史记录中的违约概率相对较高。

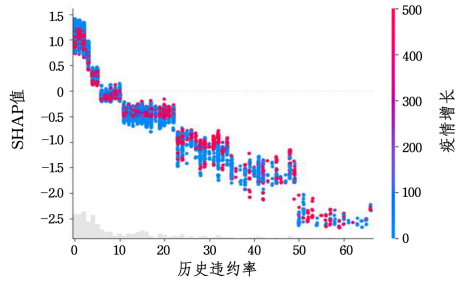

图 4 历史违约率和疫情增长的 SHAP 值散点图

Fig. 4 Scatter chart of SHAP value of historical default rate and epidemic growth

3.4.4 模型决策原因解释

为了辅助电网供应链业务人员深入理解模型决策背后的原因，图 5 显示了 20 个观测特征对 20 个实例的决策过程。

不同于整个数据集的重要性排序，决策的排序是根据观察结果计算出的重要性以降序排列，每个样本的预测结果都用彩色线表示。每个特征的 SHAP 值都添加到模型的基值中，基值中线的偏移量显示了每个特征对整体预测做出的贡

献，以及一系列特征与模型交互的累积效果。同时，决策图还可以辅助业务人员识别异常值和典型的预测路径，如图中虚线所示，其起点相对于其他偏离程度更高，可进一步探索异常的原因，以在后续的优化中识别和纠正模型的偏差。

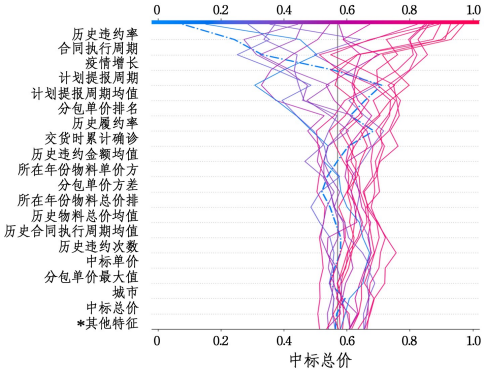

图 5 预测模型决策图

Fig. 5 Model decision diagram

3.5 多种模型对比实验结果

上述实验过程仅基于 XGBoost 算法。为了进一步验证 XGBoost 相比于其他机器学习模型在供应商履约风险中的优越性和可用性，选取 3 种机器学习领域的主流模型：*K*-临近算法（*K*-Nearest Neighbor，KNN）、梯度提升算法（Gradient Boosting Decision Tree，GBDT）以及 AdaBoost 算法，采用相同的训练集训练、测试集测试，得到的对比实验结果如表 6 所列。由表 6 可知，基于 XGBoost 模型的错误率、准确率、精确率、F_1 评分、AUC 值、KS 值均优于其他 3 种算法，充分展现了 XGBoost 的性能优势。特别地，基于 XGBoost 的预测模型取得了高达 45.38%的 KS 值，表明该模型具备较强的风险区分能力，而其他 3 种模型在供应商风险预测中的可用性较低。

表 6 多种模型的性能对比

Table 6 Performance comparison of multiple models

（%）

指标	模型名称			
	XGBoost	KNN	GBDT	AdaBoost
Error	0.07	0.10	0.10	0.10
Accuracy	93.05	89.91	90.46	90.39
Precision	94.45	91.74	90.51	90.49
Recall	99.01	97.60	99.89	99.83
F_1-*score*	96.22	94.58	94.97	94.93
AUC	86	58.68	52.13	51.98
KS	45.38	17.36	4.27	3.97

4 模型应用示例

为了将上文得到的预测模型进一步应用到电网供应链实际业务中，我们将模型输出结果转化为更直观的履约风险概率。下面介绍输入参数、输出参数以及应用示例。

4.1 输入参数

输入参数具体如表 7 所列，该输入参数为模型预测所需原始数据，由相关业务人员按照数据形式中约定或符合模型要求的形式输入到模型中。模型内部执行时会通过特征工程将这些原始数据转化为 49 个特征，自动计算各个特征值，再进行风险预测，最后得到履约概率。如需进一步分析，可取 SHAP 值，研判各个特征对该样本的贡献。

表 7 输入参数说明

Table 7 Input parameter description

输入参数	含义
中标单价	物资唯一关联的中标合同单价
中标中价	物资唯一关联的中标合同总价
类别	是否违约
物资唯一码	一个物资供应记录的唯一编码
物料编码	物料的最小类别
物料组（物料小类）	物料的次小类别
供应商编码	供应商的唯一编码
是否价格联动	供应合同是否为协议履约合同
分标编号	物料所在的分标编号
中标单价	中标合同的物料单价
中标总价	中标合同的物料总价
合同交货期	合同中约定的交货期
实际交货期	实际交货日期
采购任务接受时间	订单的接受日期
合同签订日期	签订合同的日期
计划提报日期	提报需求计划的日期
原材料类别	的原材料类别
原材料价格	原材料的全部历史价格，按月统计，列表形式
是否抽检	物资唯一码是否有抽检记录
抽检结果是否	物资唯一码抽检记录是否合格
注册资本	供应商编码关联的公司注册资本
城市	供应商公司注册地所在城市
供应商历史不良行为处罚次数	供应商编码-物料编码在供应商编码关联的不良行为处罚记录中的次数
供应商历史不良行为处罚期限总和	供应商编码-物料编码在供应商编码关联的不良行为处罚记录中的处罚期限的总和
司法协助	供应商编码关联公司司法协助次数
土地抵押面积	供应商编码关联的公司所有土地抵押记录的面积总和
欠税余额	供应商编码关联的公司欠税的余额
法律诉讼总数	供应商编码关联的公司关联的法律诉讼总数
经营异常数次	供应商编码关联的公司关联的经营异常次数
新冠疫情确诊增长	供应商关联公司所在城市在合同签订日期到系统当前时间的新冠肺炎新增数
当前新冠疫情累计确诊	供应商关联公司所在城市 在系统当前时间的新冠肺炎新增数

4.2 输出参数

供应商履约风险模型的输出参数如表 8 所列。

表 8 输出参数说明

Table 8 Output parameter description

输出参数	含义
proba	该物料供应履约的概率
values	每个特征的贡献值

以输入某一样本为例，截取输出结果，如图 6 所示。

```
proba=0.98276514
.base_values = 3.234327
.values =
array([ 2.82065481e-01, -2.06015542e-01,  3.67328942e-01,  1.72055721e-01,
       -1.82834715e-01, -1.29118934e-01,  1.21758327e-01,  1.53551891e-01,
       -1.51042193e-01, -6.93008974e-02, -3.70713137e-02, -1.24451697e-01,
        4.49693948e-02, -1.02398992e-01, -5.90600446e-03,  1.62056889e-02,
       -1.18965404e-02,  7.65076652e-03,  2.05129776e-02,  8.00902992e-02,
        4.18630838e-02, -2.19472498e-02,  7.65329301e-02, -6.22982644e-02,
        2.22731814e-01,  3.28902826e-02,  2.24365175e-01,  1.33274235e-02,
        1.66629571e-02,  6.60977364e-02,  5.01567200e-02, -2.20761318e-02,
        0.00000000e+00, -3.03236358e-02,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  2.46421580e-04,  0.00000000e+00,  0.00000000e+00,
       -3.96758132e-03,  4.78225462e-02,  0.00000000e+00, -1.00334674e-01,
        0.00000000e+00,  1.22849541e-02,  0.00000000e+00,  5.09752259e-02,
        0.00000000e+00, -2.49904506e-02, -2.70620454e-02], dtype=float32)
```

图 6 模型应用输出结果

Fig. 6 Outputs of model application

其中，*proba*＝0.982 765 14 表示该样本中供应商履约的概率为0.982 765 14。*base_values*＝3.234 327 是 SHAP 平均值，为模型各特征的综合平均表现，若要进一步分析当前影响该供应商履约的主要特征，则可以 base_values 为基准，比较其 SHAP 贡献值大小。

结束语 本文提出了基于 XGBoost 的电网供应商履约风险预测模型。针对履约过程涉及环节众多、影响履约因素错综复杂、风险特征维度丰富等问题，基于业务数据分析通过特征工程构建了 191 个风险特征，构造了本文研究所需数据集。在模型优化前后，分别进行了两次模型训练，实验结果表明，优化后模型的准确性、精确性分别高达 93.05% 和 94.45%。通过 SHAP 值方法对特征及模型结果进行解释，进而指导实际业务应用。

参 考 文 献

[1] WU Y K. Risk management of power material procurement in power grid enterprises [J]. China Logistics & Purchasing, 2019 (24): 118-119.

[2] XIE H Y, ZHONG F L, QIN M, et al. Study on the early-warning and control mechanism of the contract performance risk of power grid project materials [J]. Guangxi Electric Power, 2020 (3): 21-25.

[3] SHANG H, LEI M, MA H C, et al. Research on big data application planning method of power supply chain [J]. Electric Power, 2017, 50(6): 69-74, 94.

[4] RU L J. Study on the early-warning and control mechanism of the contract performance risk of power grid project materials [J]. Science and Technology & Innovation, 2021(15): 112-113.

[5] LI Y W, ZHANG Y, ZHANG X, et al. Design and research of risk early warning scheme for power grid material suppliers [J]. Encyclopedia Form, 2018(17): 437.

[6] QU H L, XU K, JIN N, et al. Research on knowledge graph of power material supplier evaluation based on multi data sources [J]. Smart Grid (Hans), 2020(2): 46-53.

[7] CHEN F, CHU B J. Research on knowledge atlas of power material supplier evaluation based on multiple data sources [J]. Finance & Accounting, 2020(17): 65-68.

[8] SHI L, CUI Z, YANG F, et al. The problems and solutions of power supply suppliers' performance in the post-epidemic era [J]. Electric Power Equipment Management, 2021(4): 141-143.

[9] REN M J, JIN G Q, WANG X W, et al. Microblog Popularity Prediction Algorithm Based on XGBoost [J]. Data Acquisition and Processing, 2022, 37(2): 383-395.

[10] ZHUANG J Y, YANG G H, ZHENG H F, et al. CNN-LSTM-XGBoost short-term power load forecasting method based on multi-model fusion [J]. Electric Power, 2021, 54(5): 46-55.

[11] ZHU J X, ZOU X S, XIONG W, et al. Short-Term Power Load Forecasting Based on Prophet and XGBoost Mixed Model [J]. Modern Electric Power, 2021, 38(3): 325-331.

[12] LEI X N, LIN L F, XIAO B Q, et al. Re-exploration of default characteristics of small and micro enterprises: machine learning model based on SHAP interpretation method [J]. Chinese Journal of Management Science, 2021, 27: 1-13.

[13] LIAO B, WANG Z N, LI M, et al. Integrating XGBoost and SHAP model for football player value prediction and characteristic analysis [J]. Computer Science, 2022, 49(12): 195-204.

[14] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system [C] // Proceedings of the 22nd Acmsigkdd International Conference on Knowledge Discovery and Data Mining. 2016: 785-794.

[15] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions [C] // Advances in Neural Information Processing Systems. 2017: 4765-4774.

LI Jinxia, born in 1989, master. Her main research interest is supply chain operation management.

BIAN Huaxing, born in 1989, master. His main research interest is supply chain operation management.