

大数据分析对企业信用风险评估的影响

——基于银行信贷数据的实证研究

戚莹超, 张 珍*

甘肃农业大学管理学院, 甘肃 兰州

收稿日期: 2025年3月26日; 录用日期: 2025年5月30日; 发布日期: 2025年6月9日

摘 要

数字经济时代, 金融科技快速发展推动企业信用风险评估模式的深刻变革。银行信贷数据蕴含丰富的风险信息, 通过大数据分析技术, 可以突破传统风险评估方法的局限性。该研究以某大型商业银行信贷数据为基础, 构建多维度数据分析模型, 深入挖掘企业信用风险的内在机理。运用描述性统计与关联性分析以及概率模型, 系统性地探索大数据分析对信用风险识别、预测与管理的实践价值。实证分析表明, 大数据分析显著提升了风险评估的准确性与效率, 能够精准刻画企业信用风险特征, 为金融机构信贷决策提供有力支撑。研究成果不仅丰富了企业信用风险评估的理论内涵, 还为金融机构运用大数据技术优化风险管理提供了实践指导。

关键词

大数据分析, 企业信用风险, 信贷数据, 风险评估, 金融科技

The Impact of Big Data Analysis on Enterprise Credit Risk Assessment

—An Empirical Study Based on Bank Loan Data

Yingchao Qi, Zhen Zhang*

School of Management, Gansu Agricultural University, Lanzhou Gansu

Received: Mar. 26th, 2025; accepted: May 30th, 2025; published: Jun. 9th, 2025

Abstract

In the digital economy era, the rapid development of financial technology drives profound transfor-

*通讯作者。

文章引用: 戚莹超, 张珍. 大数据分析对企业信用风险评估的影响[J]. 运筹与模糊学, 2025, 15(3): 146-150.

DOI: 10.12677/orf.2025.153148

mation in enterprise credit risk assessment models. Bank loan data contains rich risk information, and through big data analysis technology, traditional risk assessment methods' limitations can be broken through. This study, based on a large commercial bank's loan data, constructs a multi-dimensional data analysis model to deeply explore the intrinsic mechanism of enterprise credit risk. By employing descriptive statistics, correlation analysis, and probability models, the study systematically explores the practical value of big data analysis in credit risk identification, prediction, and management. Empirical analysis demonstrates that big data analysis significantly improves risk assessment accuracy and efficiency, enabling precise portrayal of enterprise credit risk characteristics and providing strong support for financial institutions' lending decisions. The research results not only enrich the theoretical connotation of enterprise credit risk assessment but also offer practical guidance for financial institutions to optimize risk management using big data technology.

Keywords

Big Data Analysis, Enterprise Credit Risk, Loan Data, Risk Assessment, Financial Technology

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

面对经济环境的复杂多变, 传统信用风险评估因信息处理能力受限, 难以应对市场快速变化。大数据分析技术应运而生, 为企业信用风险评估提供全新技术范式。通过整合多源异构数据以及先进算法, 大数据分析突破了传统单一数据源局限, 实现对企业信用风险的立体化与智能化评估。机器学习以及人工智能的引入, 使风险识别从经验性判断转向数据驱动的精准预测, 不仅提高了评估准确性, 还显著缩短了风险预警响应周期, 为金融机构构建更加动态与全面的信用风险管理体系创造了可能。

2. 文献综述

国内外学者对企业信用风险评估进行了深入研究。国际上, Kevin Koharki 和 Luke Watson (2025)发表在 *Journal of Corporate Finance* [1]的研究指出, 完善的风险管理机制能促使企业承担更多具有正净现值的项目, 从而降低信用风险。He Huang 和 Yancheng Qiu (2025)在 *Economics Letters* [2]发表的研究通过分析美国公司债券交易数据, 发现地缘政治风险与信用利差呈显著负相关, 这一发现不能由宏观经济因素或不确定性渠道解释。在国内研究方面, 余楷文[3]等(2024)基于银行逐笔信贷数据研究表明, 数字社会信用平台的建设显著降低了信贷违约风险; 王启瑞[4] (2024)发现数字化转型通过改变企业价值创造方式和信息传递模式, 强化了企业还款能力; 张雷[5] (2023)提出了基于改进梯度下降和集成学习的信用风险评估模型。这些研究虽已取得重要进展, 但在多源异构数据融合和深度学习模型应用方面仍有待深入。

3. 大数据分析与企业信用风险评估的理论基础

大数据分析与企业信用风险评估的理论基础源于金融科技发展和风险管理模式变革, 在数字经济时代, 传统评估方法因信息处理能力有限而显现瓶颈, 大数据分析为风险管理提供了新的技术范式, 通过整合多维度异构数据构建智能化预测模型, 突破了传统单一数据源和静态评估的局限, 大数据分析实现了对企业信用风险的动态监测和精准画像, 提升了风险预警的时效性和准确性, 有效降低了信息不对称程度, 推动风险管控向科学化和智能化转型。

4. 银行信贷数据的大数据分析方法

4.1. 信贷数据样本选取与数据预处理

研究选取中国东部某省份六个地级市的银行企业贷款追踪记录作为样本。数据库涵盖 103,308 家借款企业, 以小型企业为主, 占比 89.4%。行业分布以制造业为主(39.9%), 所有制类型以有限责任公司为主(65.7%)。数据预处理采用 MICE 方法处理缺失值, 针对不同类型变量采用相应的插补模型: 连续变量使用预测均值匹配, 分类变量使用逻辑回归。通过比较插补前后的数据分布, 验证了处理结果的合理性。异常值处理采用 3σ 原则, 对超出均值 ± 3 个标准差的数据点进行审查, 数据完整性得到显著提升, 缺失率从 15.3% 降低至 0.5% 以下。

数据标准化公式为:

$$x'_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

其中, x'_i 为标准化后的数值, x_i 为原始数值, μ 为均值, σ 为标准差。

经过处理后形成包含企业基本信息、财务指标、信用历史等多维度的结构化数据集。在 5,370,263 条信贷追踪记录中, 贷款违约率均值为 0.082, 标准差为 0.275, 表明样本内不同贷款的违约率存在较大差异。系统化的数据处理为深入探索大数据分析对企业信用风险评估的影响提供了可靠基础。

4.2. 大数据分析技术路径设计

大数据分析技术路径综合运用多种先进算法, 构建智能化风险评估框架。特征提取采用主成分分析 (PCA) 方法, 降维的目标函数可以表示为:

$$\max \{\lambda_1, \lambda_2, \dots, \lambda_k\} \quad (2)$$

其中 λ_i 表示特征值, 通过最大化主成分的方差贡献率, 筛选关键风险指标。

机器学习算法集成采用如下参数设置: 随机森林模型: 树的数量 500, 最大深度 10, 特征选择采用基尼系数, 支持向量机: 采用 RBF 核函数, 惩罚参数 $C = 1.0$, 核函数参数 $\gamma = 0.01$, 梯度提升树: 基于 XGBoost 实现, 学习率 0.1, 最大深度 6, 子样本比例 0.8, 深度学习网络: 三层结构, 隐藏层神经元数量分别为 128、64、32, dropout 率 0.3。

4.3. 信用风险评估模型构建

信用风险评估模型构建基于熵权法确定指标权重, 熵值计算公式为:

$$H_j = -k \sum (p_{ij} * \ln(p_{ij})) \quad (3)$$

其中, H_j 为指标熵值, p_{ij} 为标准化后的指标概率, k 为常数。

通过最小熵原理, 客观反映指标的信息量以及重要性。风险评估模型综合运用 logistic 回归以及决策树算法, 构建概率预测模型:

$$P(y) = \frac{1}{1 + e^{(-z)}} \quad (4)$$

其中, $P(y)$ 为信用风险发生概率, z 为线性组合的预测变量。

模型通过 ROC 曲线以及 K-S 统计量评估预测性能, 建立动态风险实时监测机制。最终模型能够对企业信用风险进行精准画像, 为银行信贷决策提供数据驱动的智能支持。

5. 大数据分析对企业信用风险评估的实证影响

5.1. 大数据分析对风险识别能力的提升

大数据分析通过多维度与深层次的数据挖掘,显著增强了企业信用风险识别的全面性以及精准性(表1)。传统风险识别方法受限于单一维度数据以及静态分析,而大数据分析突破了这一技术瓶颈[4]。通过整合财务指标与经营数据以及市场信息以及外部评价等多源异构数据,构建全方位的企业信用画像。机器学习算法能够捕捉复杂的非线性关系,识别潜在的风险信号以及关联模式,大幅提高风险识别的敏感度。聚类分析以及异常检测技术能够精准定位高风险企业群体,实现风险的早期预警。数据驱动的风险识别方法不仅能够突破人工识别的局限性,还能实现对企业信用风险的动态与实时监测,为金融机构提供更加全面以及深入的风险洞察。

Table 1. Key dimensions of big data analytics for enhancing risk identification capability

表 1. 大数据分析对风险识别能力提升的关键维度

维度	传统风险识别	大数据分析风险识别	提升效果
数据来源	单一维度与静态数据	多源异构与动态实时数据	显著提高
信息覆盖度	有限与碎片化	全面与系统化	大幅增加
风险信号捕捉	人工经验判断	机器学习智能识别	精准度提升
预警时效性	滞后与被动	实时与主动	明显缩短
风险评估深度	表层与片面	深层与多维	洞察力增强
样本处理能力	少量与手工	海量与自动化	极大提升
非结构化信息利用	很少	全面分析	信息价值释放

5.2. 大数据分析对风险预测准确性的改进

大数据分析显著提升了企业信用风险预测的准确性和可靠性。传统预测模型依赖有限历史数据和单一维度分析,而大数据分析通过引入机器学习和人工智能技术,实现对风险的精准预测(表2)。多模型集成方法通过综合不同算法,显著降低单一模型偏差,提高预测整体性能。神经网络算法能够捕捉复杂的非线性关系和隐藏模式,突破传统统计模型局限。

Table 2. Comparative analysis of predictive performance of different models

表 2. 不同模型预测性能对比分析

模型类型	准确率(%)	AUC 值	F1 分数	特点说明
随机森林	92.3	0.89	0.90	集成决策树, 抗过拟合能力强
支持向量机	90.1	0.87	0.88	适合小样本, 泛化能力好
梯度提升树	93.5	0.91	0.92	准确率高, 训练速度快
集成模型	94.2	0.93	0.93	综合性能最优, 稳定性强

注: 召回率 88%, 精确率 85%, 表明模型对高风险企业具有较强识别能力。时间序列验证显示预测能力稳定, 具有较好泛化性能。

通过构建动态风险预测框架,能够及时响应企业经营环境变化,为金融机构提供精准可靠的风险预测工具。模型在不同时期的稳定表现,证明了大数据分析方法在企业信用风险评估中的实用价值。

5.3. 大数据分析对信贷决策支持的实践价值

大数据分析为银行信贷决策提供全新智能化支持方案(表 3)。相比传统依赖经验的人工判断,大数据分析实现了决策的科学化以及数据驱动。先进算法快速处理海量数据,提供实时精准的风险评估,通过动态风险评分模型调整授信额度以及利率,实现差异化信贷政策。投资组合优化算法有效平衡风险与收益,实时预警机制迅速识别潜在风险,使银行能主动调整信贷策略。这种数据驱动的决策支持系统显著提升决策效率,降低信贷风险,为银行创造可观的经济以及风险管控价值。

Table 3. Comparative table of the practical value of big data analytics for credit decision support

表 3. 大数据分析对信贷决策支持实践价值的对比表

决策维度	传统信贷决策	大数据分析支持	价值提升
决策依据	主观经验判断	数据驱动与客观分析	显著提高
决策效率	人工耗时与低效	智能化与秒级响应	大幅提升
风险定价	粗放与同质化	差异化与精准化	更加精确
授信额度	静态与固化	动态与实时调整	灵活性增强
企业画像	单一维度	多维度与全景式	洞察力深化
违约预测	经验模型	机器学习预测模型	准确率提高
资产组合	经验配置	智能优化配置	风险收益平衡
决策透明度	相对封闭	可解释与可追溯	信任度提升

6. 结语

大数据分析为企业信用风险评估带来革命性变革,标志着风险管理从经验判断走向科学决策的转型。通过精准捕捉企业经营特征以及市场环境变化,大数据分析显著降低了信息不对称程度,优化了风险定价机制。金融机构借助先进的数据分析技术,可以实现对企业信用风险的动态与全面评估,有效提升风险管理效能。随着人工智能技术的持续迭代,企业信用风险评估将呈现更加科学与高效的发展态势,推动金融科技在风险管理领域的创新发展。

参考文献

- [1] Koharki, K. and Watson, L. (2025) Risk-Taking Incentives and Firm Credit Risk. *Journal of Corporate Finance*, **91**, Article ID: 102738. <https://doi.org/10.1016/j.jcorpfin.2025.102738>
- [2] Huang, H. and Qiu, Y. (2025) Does Geopolitical Risk Raise or Lower Corporate Credit Spreads? *Economics Letters*, **247**, Article ID: 112201. <https://doi.org/10.1016/j.econlet.2025.112201>
- [3] 余楷文, 申宇, 赵绍阳. 大数据对银行信贷行为的影响——来自数字社会信用平台的证据[J]. *经济研究*, 2024, 59(3): 147-165.
- [4] 王启瑞. 数字化转型对企业信用风险的影响分析[D]: [硕士学位论文]. 沈阳: 沈阳工业大学, 2024.
- [5] 张雷. 多场景下中小微企业信用风险量化与信贷决策研究[D]: [博士学位论文]. 重庆: 重庆交通大学, 2023.