



从传统分析到机器学习：债券市场信用风险识别的量化探索

技术研究部 | 徐楫淼

2026 年 9 月 29 日

摘要：

随着国内债券市场持续扩容与刚性兑付打破，信用风险呈现出更加复杂化、动态化的特征，传统依赖低频财务数据和专家经验的分析方法难以满足动态监测与前瞻预警需求，信用风险识别正加速向多源数据驱动的量化分析转型。本文围绕企业信用风险识别，系统梳理机器学习分类模型的基本原理及应用特点，并介绍特征工程、样本划分、超参数优化与模型评价等建模环节。以决策树、随机森林和 LightGBM 为代表的树模型，能够有效融合宏观、行业及企业基本面等多维信息，捕捉风险因素间的非线性关系与交互效应，其中 LightGBM 在高维结构化数据处理和计算效率方面优势突出。然而，机器学习在信用风险识别中的应用仍面临样本不平衡、过拟合以及模型可解释性有限等问题。未来，信用风险量化研究将进一步向多维特征体系、动态周期预测以及可解释机器学习等方向发展，推动传统信用研究与量化分析深度融合，为债券市场信用风险管理提供更加及时、系统的支持。

关键词：风险管理；信用风险识别；机器学习；分类模型；LightGBM

正文：

一、信用风险识别的量化转型

（一）信用风险识别的挑战

伴随着供给侧结构性改革深化、宏观经济增速换挡以及融资监管政策的收紧，企业信用风险的表现形式更加复杂，出险主体逐步从个别经营陷入困境的民营企业向地产周期、产能过剩行业、地方城投尾部主体以及政策敏感行业等领域扩散，风险呈现周期性、行业分化和区域分化特征。宏观经济、利率水平、流动性状况和信用周期等因素通过企业收入、成本、融资能力及资产价格等渠道传导至微观主体，不同行业下

企业受到的影响也存在显著差异，部分风险还可能通过产业链、担保链和融资链等渠道进一步传导，使单一企业风险演变为关联性风险。

在此背景下，企业信用风险已难以通过单一指标或少数核心指标进行充分刻画，其形成和演化往往是宏观环境、行业周期与企业基本面共同作用的结果，并具有较强的非线性和动态变化特征。传统信用研究主要依赖财务报表、信用评级等信息，虽然财务指标具有明确的经济含义和较强的可解释性，但财务数据通常按季度、半年度或年度披露，存在一定的信息时滞，且其对行业景气、市场环境和融资条件等外部变化的覆盖也相对有限。因此，面对信用风险动态化、复杂化的发展趋势，信用风险研究需要进一步拓展信息来源，将财务、市场、行业和宏观等多维信息纳入分析框架，并引入更加适合处理海量数据和复杂关系的量化方法。

（二）机器学习的引入

随着金融数据规模持续扩大以及人工智能技术不断发展，机器学习逐渐应用于金融市场分析和风险管理等领域。相较于传统统计方法主要依赖研究人员预先设定变量关系，机器学习能够通过算法从历史数据中学习特征与目标变量之间的统计关系，为复杂金融问题提供新的量化分析路径。

对于企业信用风险预测而言，机器学习的应用价值主要体现在数据处理能力和非线性关系识别能力两个方面。首先，机器学习能够拓展信用风险研究的数据维度，将信息从传统财务报表进一步扩展至宏观经济数据、行业数据、债券市场交易数据以及企业公告、新闻舆情等多维信息。其将企业基本面与外部环境信息进行综合分析，可以构建更加丰富的风险特征体系，相较于依赖少数核心指标进行判断，这种多维信息融合能够从更广泛的角度刻画企业信用状况及其变化趋势。其次，机器学习能够识别传统分析方法较难刻画的非线性关系和特征交互。企业信用风险度量通常并非多个指标简单线性叠加的结果，同一指标在不同区间内对风险的边际影响可能存在明显差异，不同风险因素之间也可能存在复杂的交互关系。以决策树、随机森林和梯度提升树为代表的机器学习算法，可以通过树结构对特征空间进行划分，并进一步通过集成学习识别变量之间的非线性关系和交互效应。

机器学习的引入可以拓展数据维度、识别复杂关系和提升计算效率，对传统信用分析形成有效补充。机器学习算法将分散于不同数据源中的风险信息转化为可计算的特征，并利用历史数据学习其中潜在的风险规律，从而为企业信用风险的识别和前瞻性预警提供新的研究工具。

二、机器学习分类模型

（一）分类模型的基本逻辑

企业信用风险识别本质上是一个典型的分类问题，即对于给定的企业及其历史经营、财务和外部环境信息，模型需要判断该企业在未来一定时期内是否可能进入风险状态。实践中我们可以将企业未来是否发生违约、评级下调或其他预设风险事件作为预测目标，并将其转化为风险和非风险两个类别。与传统信用分析主要依赖研究人员对企业经营状况进行综合判断不同，机器学习分类模型的基本思路，是利用历史样本中已经观察到的企业特征和风险结果，让算法从数据中学习二者之间的关系，并将学习到的规律应用于新的企业样本。

从形式上看，模型可以表示为： $P(Y = 1|X) = f(X)$ ；其中， X 代表企业在预测时点能够获得的各类信息，包括财务指标、市场指标、行业指标和宏观指标等； Y 代表未来是否发生预设的风险事件； $f(X)$ 则是机器学习模型根据历史数据学习得到的映射关系。对于新的企业样本，模型可以根据其特征计算相应的风险概率或风险评分，再根据设定的分类阈值判断其是否属于风险主体。

（二）模型的介绍与比较

1. 逻辑回归

逻辑回归是金融风险建模中较为经典的分类方法之一，其核心思想是通过多个解释变量对事件发生概率进行建模，并将预测结果转换为 0 至 1 之间的概率值。逻辑回归具有结构简单、计算效率较高和解释性较强等特点，模型可以通过变量系数反映不同因素与风险发生概率之间的关系。但逻辑回归通常需要对变量与风险之间的关系做出一定形式上的假设，其对复杂非线性关系和变量交互效应的刻画能力相对有限。当企业信用风险受到多个因素共同影响，且不同指标在不同区间呈现出明显不同的风险含义时，单纯依靠线性形式可能难以充分提取数据中的有效信息。

2. 决策树

决策树模型的基本思想是通过一系列条件判断逐步划分样本，即模型可以根据某一财务指标是否超过特定阈值，将企业划分为不同子群体，再继续利用其他变量进行划分，最终形成不同的分类结果。决策树最大的特点是结构直观，相比部分复杂模型具有较好的可解释性。同时，决策树能够处理变量之间的非线性关系，也不需要事先假设变量与风险之间存在某种特定的函数形式。但如果单棵决策树的深度过大，模型

容易对训练样本进行过度划分，从而学习到样本中的噪声和偶然特征，产生过拟合。因此，在实际应用中，通常需要通过限制树深度、叶节点数量等方式控制模型复杂度。

3. 随机森林

随机森林是在决策树基础上发展起来的集成学习方法，其基本思路不是建立一棵决策树，而是通过随机抽取样本和特征，构建多棵具有一定差异的决策树，再综合多棵树的预测结果得到最终分类结果。与单棵决策树相比，随机森林能够降低单一决策树对特定训练样本的依赖，在一定程度上提高模型的稳定性和泛化能力。同时，其能够处理较多的特征变量，并可以通过特征重要性等方法辅助研究人员识别影响模型预测的重要因素。不过，随机森林需要训练大量决策树，在模型规模较大时计算成本相对较高；且与单棵决策树相比，其模型结构更加复杂，整体解释难度也有所增加。

4. 梯度提升树

梯度提升树（GBDT）同样属于树模型的集成方法，但与随机森林同时构建多棵独立的决策树不同，梯度提升树采用逐步迭代的方式建立模型。模型首先建立一棵基础决策树，然后根据当前模型的预测误差建立下一棵树，对前一轮模型进行修正。经过多轮迭代后，多棵决策树共同构成最终模型。通过这种方式，模型能够不断学习前一轮预测中尚未被充分解释的信息，从而逐步提高预测能力。GBDT 能够较好地处理非线性关系和变量之间的交互效应，因此在结构化数据的分类和预测任务中得到广泛应用。基于这一思路，又发展出了 XGBoost、LightGBM 等具有代表性的梯度提升算法。

其中，LightGBM 针对大规模数据的计算效率进行了优化，通过直方图算法、叶子节点生长策略等方式提高模型训练速度，并降低计算资源需求。在金融研究中，企业财务数据、市场数据和宏观数据通常具有较强的结构化特征，同时变量数量较多、变量之间可能存在复杂的非线性关系，因此 LightGBM 在这类场景下具有较好的适用性。

表 1 不同机器学习分类模型的比较

模型	主要特点	优势	局限
逻辑回归	线性概率模型	结构简单、解释性强	非线性关系处理能力差
决策树	树结构、逐层划分	直观、解释性强	易过拟合
随机森林	多棵决策树集成	稳定性较高、抗过拟合能力较强	模型规模较大、解释难度增加
LightGBM	高效梯度提升树	计算效率高、适合高维数据	解释难度较高、参数优化要求较高

数据来源：根据公开资料整理

三、机器学习在企业信用风险识别中的应用

前文主要介绍了机器学习分类模型的基本逻辑及不同模型的特点。从实际应用来看，将机器学习用于企业信用风险识别，并不是简单地将企业数据输入模型进行训练，而是需要经过特征构建、样本划分、模型训练、参数优化和效果评价等一系列环节。其中，数据质量和特征体系是模型预测的基础，合理的样本划分和参数设置则直接关系到模型的泛化能力，模型评价则进一步决定了其风险识别能力能否得到客观检验。同时，信用风险事件属于低频事件，风险样本与正常样本之间可能存在明显的不平衡，这也对模型训练和评价提出了更高要求。基于上述考虑，企业信用风险识别的机器学习建模可以从特征工程、样本划分、超参数优化和模型评价四个方面展开。

（一）特征工程

特征工程是机器学习建模的重要基础，其核心是将原始数据转化为能够有效反映企业信用状况及其变化的模型输入变量。对于企业信用风险识别而言，特征工程不仅需要关注变量的统计预测能力，还需要充分考虑变量的经济含义、数据可得性和时间有效性。

从数据来源看，企业信用风险具有多维度影响因素，因此可以从宏观经济、行业环境、企业基本面等方面构建特征体系。宏观特征用于反映企业所处的外部经济和融资环境，包括经济增长、利率水平、流动性状况和信用环境等变量；行业特征主要用于刻画企业所处行业的景气程度及其相对位置，包括行业增长率、行业价格变化、行业盈利状况以及企业相对于行业整体水平的表现等；企业基本面特征则主要反映企业自身经营和财务状况，包括盈利能力、偿债能力、资本结构、现金流状况、资产质量以及债务规模等。

在此基础上，还可以进一步构造变化率、趋势项、波动率、历史分位数和行业相对值等衍生特征。相比单纯使用某一时点的绝对水平，变化类特征能够进一步刻画企业信用状况的边际变化，从而为风险识别提供更多信息。此外，特征数量并非越多越好。过多的冗余变量和噪声变量可能增加模型复杂度，并提高过拟合风险。因此，在特征工程过程中，需要综合运用相关性分析、统计检验以及模型特征重要性等方法，对候选变量进行筛选，在保证信息覆盖面的同时提高特征体系的有效性和稳定性。

（二）训练集、验证集与测试集

完成特征构建后，需要将历史样本划分为训练集、验证集和测试集，来分别承担

模型学习、参数选择和样本外评价的功能。其中训练集用于模型学习，即通过历史样本建立特征与风险标签之间的映射关系；验证集用于模型选择和超参数优化，用于判断不同模型或参数组合在未知样本上的表现；测试集则用于最终评价模型的样本外预测能力。

这种划分方式的核心目的是避免模型评价停留在训练样本内部，如果模型在训练数据上的表现非常好，并不能说明模型真正具有较强的预测能力，模型可能只是记住了训练样本中的特征，而没有学习到能够推广至新样本的普遍规律。因此，只有在未参与模型训练和参数选择的数据上取得稳定表现，才能说明模型具有较好的泛化能力。

（三）超参数优化

超参数是在模型训练开始之前设置的配置变量，用来控制整个学习过程。以LightGBM等梯度提升树模型为例，学习率、最大树深度、叶子节点数量以及特征和样本采样比例等均属于重要超参数。不同参数组合会对模型的拟合能力、复杂程度及泛化性能产生显著影响。当模型结构过于简单或参数设置过于保守时，模型可能无法充分提取数据中的有效信息，导致欠拟合；反之，如果模型复杂度过高，模型可能过度学习训练样本中的噪声及偶然特征，虽然能够在训练集上取得较好的拟合效果，但在样本外数据上的预测能力反而下降，即出现过拟合。因此，超参数优化的核心并非单纯追求训练集上的拟合程度，而是在模型拟合能力与样本外泛化能力之间寻求合理平衡。常见的超参数优化方法包括网格搜索、随机搜索以及基于贝叶斯优化等方法。其中，网格搜索是对预先设定的参数组合进行穷举，通过比较不同组合的验证集表现选择最优参数；随机搜索则在给定参数空间内随机抽取参数组合，在高维参数空间下通常能够以较低的计算成本探索更多可能的参数组合。

对于信用风险识别模型而言，超参数优化不能仅追求验证集上的最高预测指标，还需要综合考虑模型在不同时间窗口中的稳定性。如果某一参数组合仅在特定历史时期表现较好，而在其他时期明显下降，则其实际应用价值可能有限。因此，实际应用中需要在保证模型预测能力的同时，对模型复杂度进行适当约束，并结合时间序列验证和样本外测试，选择能够在不同市场环境下保持相对稳定表现的参数组合。

（四）模型评价指标

模型训练完成后，需要通过一系列评价指标判断其风险识别能力。对于企业信用风险识别而言，模型的目标并非简单地提高整体预测正确率，而是尽可能准确地识别潜在风险主体，同时控制误报数量。因此，需要从不同角度评价模型的分类和区分能

力。

1. 混淆矩阵

在信用风险识别中，可以将企业是否在未来一定预测窗口内发生信用风险视为一个二分类问题。模型最终需要对每个企业给出“风险”或“非风险”的分类结果，并将模型预测结果与实际风险状态进行比较，由此形成混淆矩阵。

混淆矩阵主要包含四种情况：真正例（True Positive, TP），即模型预测企业存在风险，企业实际也发生风险；假正例（False Positive, FP），即模型预测企业存在风险，但实际并未发生风险；真负例（True Negative, TN），即模型预测企业不存在风险，实际也未发生风险；以及假负例（False Negative, FN），即模型未能识别出风险，而企业实际发生了风险。

实际应用中，这四类结果的重要性并不完全相同，模型将正常企业误判为风险企业，会产生一定的误报成本，但如果模型未能识别出真正存在风险的企业，则可能错失风险处置和提前应对的窗口。因此，在评价信用风险识别模型时，不能仅关注模型预测正确的样本数量，还需要进一步区分不同类型预测错误的影响。

2. 准确度、精确度与覆盖度

基于混淆矩阵，可以进一步计算准确率（Accuracy）、精确率（Precision）和召回率（Recall）等指标。

准确率（Accuracy）：反映模型整体预测正确的比例，即所有预测正确的样本占全部样本的比例。准确率越高，说明模型整体分类结果越准确。但在信用风险识别场景下，准确率存在一定局限。由于实际发生信用风险的企业通常占比较低，即风险样本与正常样本之间可能存在明显的类别不平衡。如果模型简单地将大部分企业判断为无风险，即使遗漏了大量真正的风险企业，也可能取得较高的准确率。因此，仅以准确率衡量模型效果，容易掩盖模型在风险识别方面的实际表现。

精确率（Precision）：反映模型预测为风险的企业中，实际发生风险的比例。精确率越高意味着模型发出的风险识别越准确，将正常企业误判为风险企业的情况越少。从实际业务角度看，较高的精确率有助于提高后续人工核查和风险处置的效率。

召回率（Recall）：反映实际发生风险的企业中，有多少被模型成功识别出来，也可以理解为模型对风险主体的覆盖能力。对于信用风险识别而言，召回率通常具有较高的重要性。如果一个模型虽然发出的预警数量不多，但能够覆盖绝大多数真正的风险主体，则其风险筛查价值相对较高。

此外，精确率和召回率之间通常存在一定的权衡关系。提高模型识别风险的敏感程度，可能增加风险识别数量，同时也可能带来更多误报；反之，如果模型设置得更加严格，虽然能够提高预警的准确程度，但也可能遗漏部分潜在风险。因此，在实际应用中，需要结合风险管理目标和后续核查能力合理确定模型的分类阈值，而不是单纯追求某一个指标的最大化。

3. ROC-AUC

除上述指标外，ROC-AUC 也是分类模型中较为常用的评价指标。与准确率、精确率和召回率通常对应某一个具体分类阈值不同，ROC-AUC 主要用于评价模型在不同阈值下区分风险企业与非风险企业的整体能力。

在信用风险模型中，模型通常并不是直接输出风险或非风险两个确定结果，而是首先给出一个介于 0 和 1 之间的风险概率或风险评分，再根据设定的阈值进行分类。ROC 曲线正是通过考察不同阈值下模型的表现，对模型整体区分能力进行评价；AUC 则表示 ROC 曲线下的面积，其取值通常介于 0 和 1 之间，数值越高，说明模型区分风险企业和非风险企业的能力越强。对于企业信用风险识别而言，ROC-AUC 的优势在于不依赖某一个具体的分类阈值，因此可以用于不同模型之间的横向比较。但同样需要注意，AUC 较高并不意味着模型一定适合直接用于实际风险识别。特别是在信用风险事件比例较低的情况下，还需要结合精确率、召回率以及不同风险阈值下的实际表现进行综合判断。

在实际应用中，模型评价最终还需要进一步回到业务目标，即模型能否在较少增加无效预警的情况下，尽可能提前识别潜在风险主体，并在不同时间区间和市场环境下保持相对稳定的表现。

四、总结与展望

随着债券市场信用风险特征日益复杂，信用风险识别正在从依赖低频财务信息和专家经验的传统模式，逐步向多维数据驱动的量化分析转变。通过整合企业财务、市场交易、行业环境及宏观经济等多维信息，机器学习能够从历史数据中挖掘风险因素之间的非线性关系和交互效应，为企业信用风险识别和预警提供更加丰富的量化依据。

从实际建模过程来看，机器学习在信用风险识别中的应用并非简单地选择一个算法并进行训练，而是涉及特征构建、样本划分、参数优化和模型评价等多个环节。其中，特征体系决定了模型能够获得的信息边界，样本划分和验证方式关系到模型能否得到客观评价，超参数优化影响模型的复杂程度与泛化能力，而评价指标则从不同角

度衡量模型的风险识别效果。同时，模型输出的风险概率或评分更多是一种风险信号，仍需要结合企业运行情况以及市场信息进行进一步分析和判断。尤其是在实际应用中，模型的可解释性、数据质量、样本偏差和过拟合等问题，都可能影响预测结果的可靠性，这也是机器学习从研究走向实际业务应用过程中需要持续解决的问题。

展望未来，机器学习在信用风险研究中的应用仍有较大的拓展空间。一方面，随着数据来源不断丰富，信用风险模型可以进一步融合债券市场、产业链、舆情及企业运行等信息，形成更加动态的风险特征体系；另一方面，模型方法也将从单一算法预测逐步向模型组合、动态更新和多周期预测等方向发展。同时，可解释机器学习等方法的发展，也有望进一步增强模型结果与传统信用研究之间的衔接。机器学习的意义并不在于以算法取代传统信用研究，而在于借助数据和模型提升研究的覆盖范围、分析效率和风险识别的前瞻性。随着数据基础和模型方法不断完善，机器学习有望与传统信用研究形成更加紧密的结合，为债券市场信用风险管理提供更加及时、系统的量化支持。

参考文献：

- [1] 周璇. 基于机器学习算法的债券违约风险预警研究[J]. 财经界, 2025, (22): 29-31. DOI: 10. 19887/j. cnki. cn11-4098/f. 2025. 22. 007.
- [2] 周荣喜, 彭航, 李欣宇, 等: 基于 XGBoost 算法的信用债违约预测模型[J]. 债券, 2019, (10): 61-68.
- [3] 胡传胜, 周志国. 基于机器学习算法的企业债券违约预警模型[J]. 长江信息通信, 2024, 37(12): 173-175+202. DOI: 10. 20153/j. issn. 2096-9759. 2024. 12. 051.
- [4] 迟雪丹. 中国信用债券违约风险度量与预警机制研究[D]. 吉林大学, 2024. DOI: 10. 27162/d. cnki. gjlin. 2024. 000372.



报告声明

本报告分析及建议所依据的信息均来源于公开资料，本公司对这些信息的合法性、真实性、准确性和完整性不作任何保证，也不保证所依据的信息和建议不会发生任何变化。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不构成任何投资建议或决策依据。投资者依据本报告提供的信息进行证券投资所造成的一切后果，本公司概不负责。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为大公国际，且不得对本报告进行有悖原意的引用、删节和修改。