

宏观环境感知的个人贷款风控系统

一、项目名称

「MacroGuard」—— 宏观环境感知的个人贷款风控系统

二、研究动机与背景

2.1 项目受众与现实意义

本项目的直接受众是**商业银行风控部门、金融科技公司的信贷决策团队、金融监管与审计机构**，以及从事金融 AI 研究的数据科学家。在更宏阔的层面，每一个申请个人贷款的普通消费者都是间接受众——他们有权知道，决定他们贷款命运的 AI 系统是否公平、透明、经得起追问。

现实意义体现在三个层面。其一，**经济价值**：美国个人贷款市场年规模超过 2000 亿美元，违约率每降低一个百分点，即可为金融机构挽回数十亿美元的坏账损失。其二，**社会价值**：传统信用评分体系将大量缺乏信用记录的“信用隐形”人群排斥在金融服务之外，而多源数据 AI 模型有望更全面、更公平地评估其还款能力。其三，**技术示范价值**：本项目构建了一条从数据融合到可解释建模、从风控策略到 LLM 交互的完整链路，为金融 AI 从实验室走向真实业务提供了可复现的参照范式。

2.2 相关工作

在信用风险建模领域，传统方法长期依赖逻辑回归和评分卡体系，以 FICO 评分为代表。这类方法可解释性强，但对非线性关系和高维交互效应的捕捉能力有限。近年来，XGBoost、LightGBM 等梯度提升树模型在 Kaggle 借贷竞赛中表现优异，但"黑箱"属性使其难以通过监管审计。

在可解释性方面，Lundberg 等人提出的 SHAP 框架为树模型的特征归因提供了理论基础，但大多数研究仅止步于特征重要性排序，未能将模型输出进一步转化为可执行的风控策略。在多源数据融合方面，现有研究多局限于单一数据集（如 Lending Club），鲜有将宏观经济指标和区域经济数据系统性地融入违约预测的工作。本项目正是在这些交叉点上进行了工程化探索。

三、问题定义与挑战

3.1 核心问题

本项目试图回答四个逐层递进的问题：

- 预测：**在宏观非平稳环境下，如何利用多源数据准确预测个人贷款的违约概率？
- 解释：**模型为什么会给出某个预测？哪些因素驱动了决策？特征是相关还是因果？
- 决策：**预测概率如何转化为可落地的风控策略？在通过率和坏账率之间如何找到最优平衡？
- 交互：**如何让不具备数据科学背景的业务人员和监管者也能理解并信任模型决策？

3.2 关键挑战

挑战一：多源异构数据融合。 三套数据——Lending Club 226 万条贷款记录（2007-2018）、FRED 宏观经济月度指标、USDA ERS 50 州的区域经济数据——在时间粒度（月/季/年）、空间粒度（州/县）和个体粒度（单笔贷款）上完全不同，需要精细的对齐策略。

挑战二：宏观非平稳性。 数据跨越了 2008 年金融危机及其后的量化宽松、加息周期。概念漂移分析显示 FICO 平均分的 PSI 值高达 3.0，表明借款人群体和信用环境发生了根本性变化。传统的随机划分训练/测试集会高估模型泛化能力。

挑战三：可解释性与预测性能的平衡。 监管机构（如美联储的 SR 11-7 指引）要求模型具备可审计性，但高性能模型往往是黑箱。需要在 SHAP、因果推断、反事实解释等层面构建多层可解释性体系。

挑战四：从预测到决策的鸿沟。 一个 AUC 0.71 的模型本身不能直接创造商业价值。需要设计动态阈值策略、压力测试框架、CECL 准备金计算和组合优化方案，将概率转化为行动。

四、解决方法

4.1 技术路线总览

系统沿五层架构展开：**数据融合层** → **智能建模层** → **可解释性层** → **风控策略层** → **LLM 智能交互层**（在第七节的数据产品中详述），由 26 个 Python 脚本串联为可一键复现的完整链路。

4.2 数据融合层

时序特征工程包含三组计算：基于借款人 FICO、利率、贷款金额构造 30/60/90 天滚动窗口均值和标准差，捕捉近期信用趋势；时间衰减特征赋予近期事件更高权重；季节和节假日二元特征捕捉周期效应。

跨源特征融合将 FRED 宏观三因子（联邦基金利率、失业率、CPI 通胀率）按贷款发放月份精确对齐，ERS 州级三因子（家庭收入中位数、贫困率、失业率）通过 `addr\_state` 连接。在此基础上构造三个经济学含义明确的交互项：`int\_rate × fed\_funds\_rate`（货币政策传导效应）、`fico\_avg × cpi\_inflation`（信用质量与通胀的交互）、`loan\_amnt × state\_unemployment\_rate`（区域经济脆弱性敞口）。

4.3 智能建模层

采用**时间序列划分**策略（2007-2016 训练，2017-2018 测试）而非随机交叉验证，以真实模拟模型在未来的泛化表现。基准模型包括 Logistic Regression（作为可解释性基线）和 XGBoost（作为性能上限）。在此基础上延伸出贝叶斯逻辑回归（量化预测不确定性）、Cox 比例风险生存分析（预测"何时违约"而非仅"是否违约"）、以及基于 Optuna 的 AutoML CASH 联合搜索（同时优化模型族、预处理、特征工程和不平衡处理策略）。

4.4 可解释性层

三层递进：**归因层**——SHAP 值量化每个特征对单笔预测的贡献；**因果层**——DID 估计利率政策变更的处理效应，工具变量法处理内生性问题；**反事实层**——回答"如果 FICO 提高 50 分，审批结果会改变吗"这类对借款人至关重要的 what-if 问题。

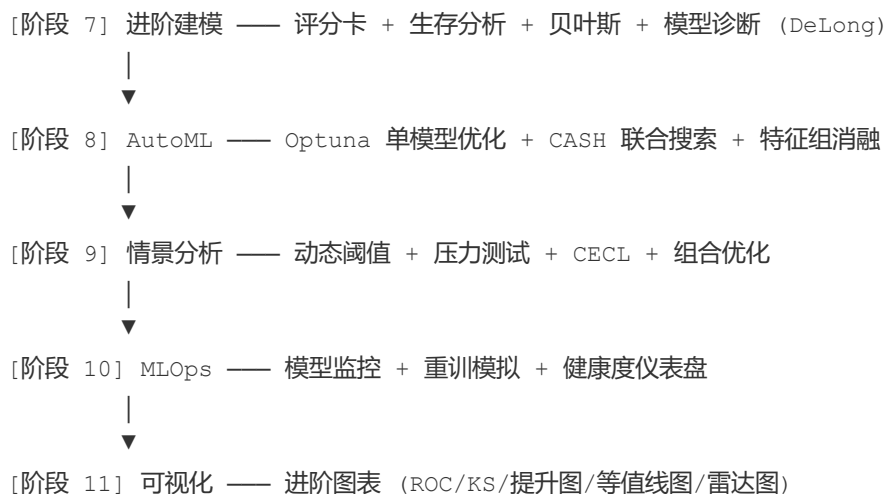
4.5 风控策略层

以利润函数 $\text{\$Profit} = \text{Approved} \times [(1 - \text{BadRate}) \times \text{InterestMargin} - \text{BadRate} \times \text{LGD}]$ 为目标，在 0.05 到 0.95 的阈值区间扫描最优决策边界。在此基础上引入**宏观状态感知**机制——用联邦基金利率的分位数将经济环境划分为正常期、观察期、压力期三种状态，动态调整审批阈值偏移量 ± 0.05 。辅以多维压力测试和 Markowitz 均值-方差组合优化，为不同风险偏好的投资组合提供有效前沿。

五、数据流水线

5.1 全链路架构





5.2 关键组件剖析

违约标签构造 (``analysis/analyze_lending_club.py``)：从 226 万条原始记录中识别已完结贷款 136 万条，其中 "Fully Paid" 标记为 0，"Charged Off / Default / Late(31-120)" 标记为 1，整体违约率 21.27%。排除 "Current / In Grace Period / Late(16-30)" 等未完结状态 89 万条。

共享数据构造器 (``common/model_data.py``)：集中式 ``build_training_sample()`` 函数，整合原始字段、衍生字段、时序特征和跨源特征，确保所有下游脚本使用一致的特征口径。按 ``issue_year`` 进行时间序列划分。

LLM 客户端 (``common/llm_client.py``)：统一的 OpenAI 兼容协议封装，从 ``env`` 自动加载凭证，支撑四个 LLM 子模块。

六、实验分析

6.1 模型性能验证

在 112 万训练集 (2007-2016) 和 24 万测试集 (2017-2018) 上, XGBoost 以 AUC 0.709、KS 0.303 显著优于 Logistic Regression (AUC 0.678、KS 0.260)。DeLong 检验的 p 值趋近于零 ($p \approx 0.000000$), 确认 XGBoost 的 AUC 提升并非随机波动。

6.2 特征消融实验

AutoML 特征组消融实验提供了关键证据。纯贷款特征 (base, 18 维) AUC = 0.6995; 加入时序特征 (with_temporal, 25 维) 后提升至 0.7057 (+0.0062); 融合全部跨源特征 (all_state_aware, 34 维) 后达到最高 AUC 0.7084 (+0.0089 vs base), Top Decile 坏账捕获率从 19.23% 提升至 20.36%。这组消融实验有力地证明了多源数据融合的价值。

6.3 宏观环境的影响

相关性分析揭示了联邦基金利率与违约率的极强关联——年度相关系数高达 +0.94, 季度相关系数 +0.84。州级层面, 家庭收入中位数与违约率的相关系数为 -0.40, 贫困率为 +0.40。即使在控制了贷款等级和利率结构之后, 贫困率的残差相关性仍保持在 0.36, 说明区域经济确实对个体违约行为存在独立的解释力。

6.4 贝叶斯建模验证

贝叶斯逻辑回归提供了预测不确定性的量化。`term\_months` 的后验均值 +0.324 (95% CI [0.279, 0.370])、`fed\_funds\_rate` 的后验均值 +0.194 (95% CI [0.105, 0.283]) 均显著非零, 从贝叶斯视角验证了跨源宏观特征对违约风险的独立贡献。

6.5 概念漂移的科学处理

PSI 分析揭示了极为显著的特征分布偏移——FICO 平均分在 2013-2016 年间的 PSI 稳定在 2.9-3.0 的超高水平，`int\_rate` 在 2009 年 PSI 达 2.18。这意味着随机交叉验证会产生严重的信息泄露和过分乐观的估计。本项目严格采用基于年份的时间序列划分（训练/测试在时间上不重叠），确保了性能评估的科学性。

七、数据产品

最终交付的是一个基于 Streamlit 的**交互式智能风控看板**，共 7 个 Tab：

Tab 1：数据概览
· 2,260,701 条记录 / 21.27% 违约率
· 数据质量报告 + 概念漂移 PSI 热力图
· 缺失值分布 / 异常值统计 / 类别不平衡分析
Tab 2：模型表现
· LR vs XGBoost 指标对比表 (AUC/KS/PR-AUC/Brier)
· Top Decile 坏账捕获率 + ROC/KS/校准/提升曲线
· DeLong 检验 + 学习曲线 + 子群体校准
Tab 3：AutoML
· 5 组特征集消融对比
· 模型家族性能排序 (LR/XGBoost/LightGBM/RF/ET)
· 业务指标：最佳阈值 / 通过率 / 坏账率 / 单笔利润
· CASH 搜索历史 + 超参数重要性
Tab 4：可解释性
· SHAP Beeswarm / Bar / Summary / Interaction
· PDP 部分依赖图
· 因果 DID / 工具变量 / 中介效应

Tab 5: 风控策略
· 19 级阈值扫描 (0.05-0.95)
· 利润最优策略: 阈值 0.15 → 通过率 21% / 坏账率 8.18%
· 状态感知动态策略: 宏观环境自适应阈值调整
Tab 6: 决策追溯
· 单笔贷款反事实解释: FICO/利率/收入逐特征扰动
· 最小改变量分析: 从拒绝到批准需要改变什么
· 完整决策审计链路
Tab 7: AI 助手 (4 个并发子 Tab)
· 自然语言问答 - Text-to-Pandas (AST 沙箱安全执行)
· RAG 检索 - TF-IDF 索引 + 证据溯源
· 决策解释 - LLM 生成自然语言审批理由
· Agent - 多工具自主调用 + 复杂综合查询

技术亮点：AI 助手底层使用 `ThreadPoolExecutor` + `st.fragment` 局部轮询，4 个子 Tab 互相独立，用户可同时发起四个长任务而互不阻塞。

完整的项目产出数据：**76 张统计表 (CSV/Markdown)**、**60 张专业图表 (PNG)**、**11 个序列化模型文件 (joblib)**、**SHAP 值矩阵**、**RAG 向量索引**，全部由 `python main.py` 一键生成。

八、经验与教训

8.1 技术阻碍与攻克

阻碍一：海量数据的单机处理。 226 万行 CSV 在 pandas 中的内存占用超过 2GB，32GB MacBook 在串联多个脚本时反复出现内存压力。解决方案是在每个脚本完成后显式 `del` 大

对象并调用垃圾回收，同时将概念漂移等计算密集型任务限制在采样数据上。另一个教训是将特征工程集中在数据加载阶段完成，避免每个下游脚本重复读取和转换。

阻碍二：自签名 SSL 证书与 Safari 兼容。 macOS 上的 Safari 默认强制 HTTPS 无法访问本地 Streamlit 的 HTTP 端口。通过 OpenSSL 生成自签名证书，并以 `--server.sslCertFile` 和 `--server.sslKeyFile` 参数启动 Streamlit 解决。这提醒我们在开发环境中也应尽早考虑部署形态。

阻碍三：Python 脚本的导入路径管理。 项目中的脚本分布于 8 个子目录，使用 `subprocess.run` 统一调度时，`cwd` 参数不足以保证所有脚本正确导入 `constant` 等顶层模块。解决方案是在 `subprocess.run` 的 `env` 中显式设置 `PYTHONPATH`。

阻碍四：因果分析中的特征列不匹配。 反事实解释器仅使用了 `NUMERIC_FEATURES` 和 `CATEGORICAL_FEATURES` 列表，但训练好的 Pipeline 期望完整的 34 维特征输入（含 `CROSS_SOURCE_NUMERIC_FEATURES`），导致 `ValueError: columns are missing`。修复方法是补全特征列表，这提醒我们特征工程的“口径一致性”是机器学习流水线中最易出错的环节。

阻碍五：matplotlib 中文字体缺失。 因果分析图表的图例包含中文文字，在 macOS 的 DejaVu Sans Mono 字体下渲染为乱码或方块。解决方案是将图表文字限定为英文，或将 `PingFang SC` 注册为默认字体。对于面向国际受众的技术报告，建议全部使用英文标注以提高可移植性。

8.2 核心经验

1. 时间序列划分是金融建模的生命线。 本项目的 PSI 分析表明特征分布存在剧烈漂移，随机交叉验证会给出误导性偏高的性能估计。严格按年份划分训练和测试集不是可选项，而是必须遵守的纪律。

2. 多源数据的价值需要特征消融来证明。 仅仅声称使用了宏观数据是不够的，需要用 base/with_temporal/with_macro/with_region/all_state_aware 的消融对比来量化每类数据源的边际贡献。

3. 从模型到策略需要设计利润函数。 AUC 本身不创造商业价值。只有当阈值扫描与利润函数、压力测试、准备金计提、组合优化串联时，AI 系统才真正具备了业务闭环能力。

4. 可解释性不止于特征重要性。 SHAP 排名是起点，不是终点。因果推断区分了相关与因果，反事实解释提供了可操作的"what-if"推理，LLM 自然语言解释让非技术人员也能理解——三层递进才是完整的可解释性体系。

5. LLM 集成要兼顾安全与用户体验。 Text-to-Pandas 的代码生成必须经过 AST 黑名单过滤（禁止 import/open/dunder 等高危关键字），RAG 回答必须标注证据来源，这些安全基线不能因为追求功能丰富性而妥协。

九、项目总结

这是一个面向个人贷款违约风险的多源数据智能风控系统。它以 Lending Club 226 万条信贷记录为核心，融合 FRED 宏观三因子和 USDA ERS 50 州经济指标，构建了 34 维特征空间，并通过时间序列划分策略（2007-2016 训练 / 2017-2018 测试）保证性能评估的科学性。

在建模层面，XGBoost 实现了 AUC 0.709、KS 0.303 的违约预测能力——这是一个在谨慎的时间序列验证下取得的、经 DeLong 检验统计显著的性能。AutoML 特征组消融实验表明，多源数据融合比纯贷款特征提升了 +0.009 AUC，Top Decile 坏账捕获率从 19.23% 提升至 20.36%。在策略层面，利润最优阈值 0.15 将审批池的坏账率从原始 21.27% 压缩至 8.18%，宏观状态感知动态策略进一步将单笔贷款损失从 -493 降至 -436，改善幅度达 13.3%。

在可解释性层面，SHAP + 因果推断 + 反事实解释的三层体系使模型不再是黑箱——每一个决策都可以被归因、被验证、被追问。在交互层面，四重 LLM 能力（自然语言问答、RAG 检索、决策解释、Agent）集成于一个 Streamlit 看板，让数据、模型和策略从工程师的终端走向了业务人员的浏览器。

这是一场关于“信任”的技术实验。当 AI 决定一个人的贷款命运时，它凭什么被信任？我们的答案是：**多源数据交叉验证，模型预测可解释，风控决策可审计，整个过程可对话。**当全球数以亿计缺乏传统信用记录的人，第一次有机会获得公平的金融服务时，这套“数据融合 + 可解释 AI + 智能交互”的技术基座，将成为那个未来的起点。

附录：核心指标速览

维度	指标	数值
数据	原始贷款记录数	2,260,701
数据	可用标签样本数	1,367,578
数据	整体违约率	21.27%
数据	覆盖年份	2007–2018
数据	外部数据源	FRED + USDA ERS (50 州, 47 季度)
最佳模型	算法	XGBoost
最佳模型	AUC	0.709
最佳模型	KS 值	0.3033
最佳模型	PR-AUC	0.4488
最佳模型	Brier 分数	0.1735
最佳模型	Top 10% 坏账捕获率	20.29%
AutoML	最优特征集	all_state_aware (34 维)
AutoML	最优特征集 AUC	0.7084
AutoML	vs Base 特征集提升	+0.0089 AUC

风控策略	最优固定阈值	0.15
风控策略	通过率 @ 最优阈值	21.11%
风控策略	坏账率 @ 最优阈值	8.18%
风控策略	单笔利润 (XGBoost)	0.006
风控策略	动态策略 vs 固定基线	亏损减少 13.3%
宏观关联	联邦基金利率 (年度)	$r = +0.94$
宏观关联	联邦基金利率 (季度)	$r = +0.84$
宏观关联	州级贫困率	$r = +0.40$
宏观关联	州级收入中位数	$r = -0.40$
可解释性	DID 效应	0.489 (OR=1.63, $p \approx 0$)
可解释性	DeLong 检验 (XGB vs LR)	$p \approx 0.000000$
可解释性	概念漂移 (FICO PSI)	高达 3.05 (2013)
CECL	总计提准备金 (5 万笔)	~\$113.7M
工程	Dashboard Tab 数	7
工程	LLM 能力数	4

工程	一键复现命令	python main.py
----	--------	----------------