

客脉守望者—咖啡馆等小型业态多因子特征工程与随机森林流失预警模型

一、项目简介

“客脉守望者”是一个面向精品咖啡馆等小微商业场景的客户流失预警系统。系统以机器学习为核心，基于随机森林分类器，结合多因子特征工程，从客户消费行为、互动记录、价值贡献等多个维度构建预测模型，精准识别存在流失风险的客户。项目通过 Flask 框架搭建 Web 可视化平台，实现数据上传、模型预测、结果展示与运营建议的一体化管理，帮助商家从“被动应对”转向“主动挽留”，提升客户留存率与经营效益。

二、项目特点：

多因子特征融合，提升预测精度：系统不仅使用 RFM 基础指标，还引入“平均消费间隔”“活动参与度”“登录活跃度”“偏好稳定性”等多维特征，全面刻画客户行为模式，显著增强模型判别能力。

随机森林算法，兼顾准确性与可解释性：采用随机森林分类器，具备强鲁棒性、抗过拟合能力，同时输出特征重要性排序，帮助商家理解“哪些行为最可能导致客户流失”，实现“可解释的 AI 决策”。

端到端系统闭环，落地性强：从原始数据清洗、特征构建、模型训练到前端展示，构建完整数据流水线，支持 CSV 批量导入与自动预测，具备实际部署与持续迭代能力。

可视化交互设计，赋能非技术用户：基于 Flask+ECharts 打造直观管理后台，通过雷达图、散点图、风险排行榜等形式，将复杂模型结果转化为可操作的商业洞察，降低使用门槛。

三、项目收获

技术层面：深入掌握机器学习全流程，包括数据预处理、特征工程、模型选型与评估、交叉验证等；熟练运用 Scikit-learn、Pandas、Flask 等 Python 技术栈，提升全栈开发能力。

业务层面：理解客户生命周期管理与精细化运营逻辑，学会从商业问题出发设计技术方案，实现“技术驱动业务”的思维转变。

综合能力：锻炼项目管理、系统设计、可视化表达与文档撰写能力，形成从“想法”到“产品”的完整闭环经验。

四、项目成果

建成一个可运行的客户流失预警系统：实现数据上传、自动建模、风险评分、可视化展示等核心功能，系统稳定运行于本地 Flask 服务。

模型性能优异：在测试集上，随机森林模型准确率达 85%以上，AUC 值超过 0.88，显著优于逻辑回归等基线模型，具备实际应用价值。

输出客户风险分级清单：系统可定期生成“高流失风险客户名单”，并附带关键行为特征分析，为商家提供精准挽留依据。

五、项目安排

1、项目总时长：项目学习与实 30 小时

2、项目执行大纲

执行大纲

课程板块	板块 课时	课时	内容主题	关键学习点或任务
第一部分：商业问题 拆解与数据基石	10	2	微观商业场景与建模目标 定义	聚焦精品咖啡馆等小微商业场景，拆解“客户流失”的核心商业问题
		2	多源数据采集与清洗实战	讲解客户数据来源，使用Pandas 读取 CSV、Excel 等格式数据；处理缺失值、异常值；统一数据格式，构建干净、规范的基础数据集
		2	RFM 模型与基础特征构建	讲解 RFM 模型的核心逻辑，使用Pandas 计算每个客户的 RFM 指标；对 RFM 指标进行分箱处理，为后续聚类与分类模型提供基础特征。
		2	多因子特征工程进阶	拓展特征维度，构建“行为活跃度特征”、“消费偏好特征”、“互动特征”
		2	特征筛选与相关性分析	使用 Pearson 相关系数分析特征与目标变量的相关性，剔除低相关特征；通过方差筛选与业务逻辑筛选，最终确定 10 个核心特征；使用 Seaborn 绘制特征相关性热力图，直观展示特征间的关联关系。
第二部分：随机森林 模型构建与优化	10	2	随机森林算法原理与选型	讲解随机森林的核心逻辑，对比决策树、逻辑回归、随机森林的优劣；使用 Scikit-learn 的 RandomForestClassifier 初始化模型，讲解关键参数的作用
		2	模型训练与基线评估	使用 train_test_split 划分训练集与测试集，采用分层抽样保持流失客户比例；训练初始随机森林模型，使用准确率、召回率、F1-score、AUC 等指标评估基线性能；绘制混淆矩阵，分析模型对“流失客户”的识别能力
		2	特征重要性分析与业务解读	绘制柱状图排序；结合业务逻辑解读特征重要性；剔除重要性低于 0.05 的冗余特征，简化模型复杂度。

		2	模型调优与过拟合控制	使用网格搜索调优关键参数，以 AUC 为优化目标；通过交叉验证评估调优后模型的稳定性，避免过拟合；对比调优前后模型性能，确定最优参数组合。
		2	模型部署与预测接口开发	使用 joblib 库保存训练好的模型，编写预测函数，输入新客户特征矩阵，输出流失概率与风险等级；测试预测接口的准确性。
第三部分：Flask 系统开发与成果展示	10	2	Flask 框架基础与项目架构	讲解 Flask 的核心组件，搭建项目基础架构；编写首页路由，渲染包含“数据上传”“风险查询”功能的主页面；使用 Flask 的 request 对象处理文件上传
		2	数据上传与模型调用模块	编写文件上传接口，使用 Pandas 读取上传的 CSV 文件，校验数据格式；调用保存的模型进行批量预测，将预测结果写入 DataFrame；使用 Flask 的 session 存储预测结果，避免重复计算。
		2	ECharts 可视化模块开发	在 templates 文件夹中创建 risk_dashboard.html 页面，引入 ECharts 库；绘制“客户风险等级分布饼图”、“核心特征重要性柱状图”；使用 Flask 的 jsonify 接口传递预测结果数据，实现动态图表渲染
		2	风险客户列表与运营建议模块	编写风险客户查询接口，按流失概率降序排列客户，展示 Top20 高风险客户的核心信息；结合特征重要性生成运营建议，在页面中以卡片形式展示
		2	系统测试与成果整合	进行端到端测试，修复界面显示异常、数据传递错误等问题
课时合计		30		

*备注：项目根据该大纲执行，如遇某些临时情况，将会再做沟通调整。

六、时间安排

项目执行

2026 年 3 月至 2026 年 8 月，项目学习与软件实现（30 小时）

七、项目费用：

成果	价格	人数	备注
客脉守望者—咖啡馆等小型业态多因子特征工程与随机森林流失预警模型		1 人	费用包含：方案设计、技术指导、程序开发。 费用不包含：可能产生的第三方收费，如网站域名费、服务器租用费、AI 工具使用费、大模型的 token 费用（预计整体费用 3 千元左右，具体根据运营使用情况而定）