

Ant-X 学业风险预警模型

案例提供部门：学生事务办公室

支持部门：西浦学习超市

1. 案例背景

(1) 旧系统的局限性

- **覆盖范围有限**：仅针对大陆本科生，未涵盖海外学生及硕博研究生群体。
- **数据治理不足**：学校数据收集范围不全面，风险预警依赖事后统计，无法提前预判（如挂科预警滞后）；跨部门数据协同不足，数据价值未充分释放。
- **校外供应商风险**：引入外部供应商开发时，存在数据离域（隐私泄露）、算法黑箱不可控等问题。

(2) 中止原因与重启背景

- **中止原因**：受疫情影响，校外供应商技术骨干流失，缺乏后续开发能力，项目一度停滞。

- **重启契机**：2025 年，学校数据治理成效显著，具备风险预警的数据基础；在两位分管校长支持下，以校内特殊科研项目方式立项，解决了预算流转与跨部门协作的历史障碍。

2. 解决方案

(1) 跨中心协作机制

组建由多方构成的跨中心工作组：

- **业务专家**：SAO 和教务处，负责需求定义与业务验证；
- **技术专家**：MITS，负责数据安全、系统部署与 API 开发；
- **模型专家**：SAT 王秋锋教授团队，负责算法研发与模型优化。

(2) 校内特殊科研立项

通过校内特殊科研项目（RDF）立项，规避校外供应商风险，确保数据安全与算法可控；建立定期项目例会制度，高效推进开发。

(3) 模型开发与功能设计

- **算法模型**：
 - **无监督学习**：采用 Local Outlier Factor（LOF）、Isolation Forest（孤立森林），支持参数调整优化异常检测；
 - **有监督学习**：包含二分类（是否存在风险）与回归（风险概率）任务，模型结构固定且精度高。

- **预警机制：**构建“预警-干预-反馈”闭环流程——风险预警任务自动派发，学生成长顾问（DA）记录干预过程并反馈模型准确性。

(4) 上线应用步骤

- Phase 1：完成学业风险模型开发（挂科概率、群体偏离度）；
- Phase 2：扩展至心理预警（短期应激异常）；
- Phase 3：规划就业困难预测；

按步骤推进，逐步实现系统上线：包含数据加密、权限管理、工单自动派发等功能。

3. 成果与效益

(1) 模型性能显著

基于 AY2023-2024 数据验证，4 个核心模型的 F1 分数（分类任务）与 R²值（回归任务）均达 80%以上，预测可靠性超过行业基准。

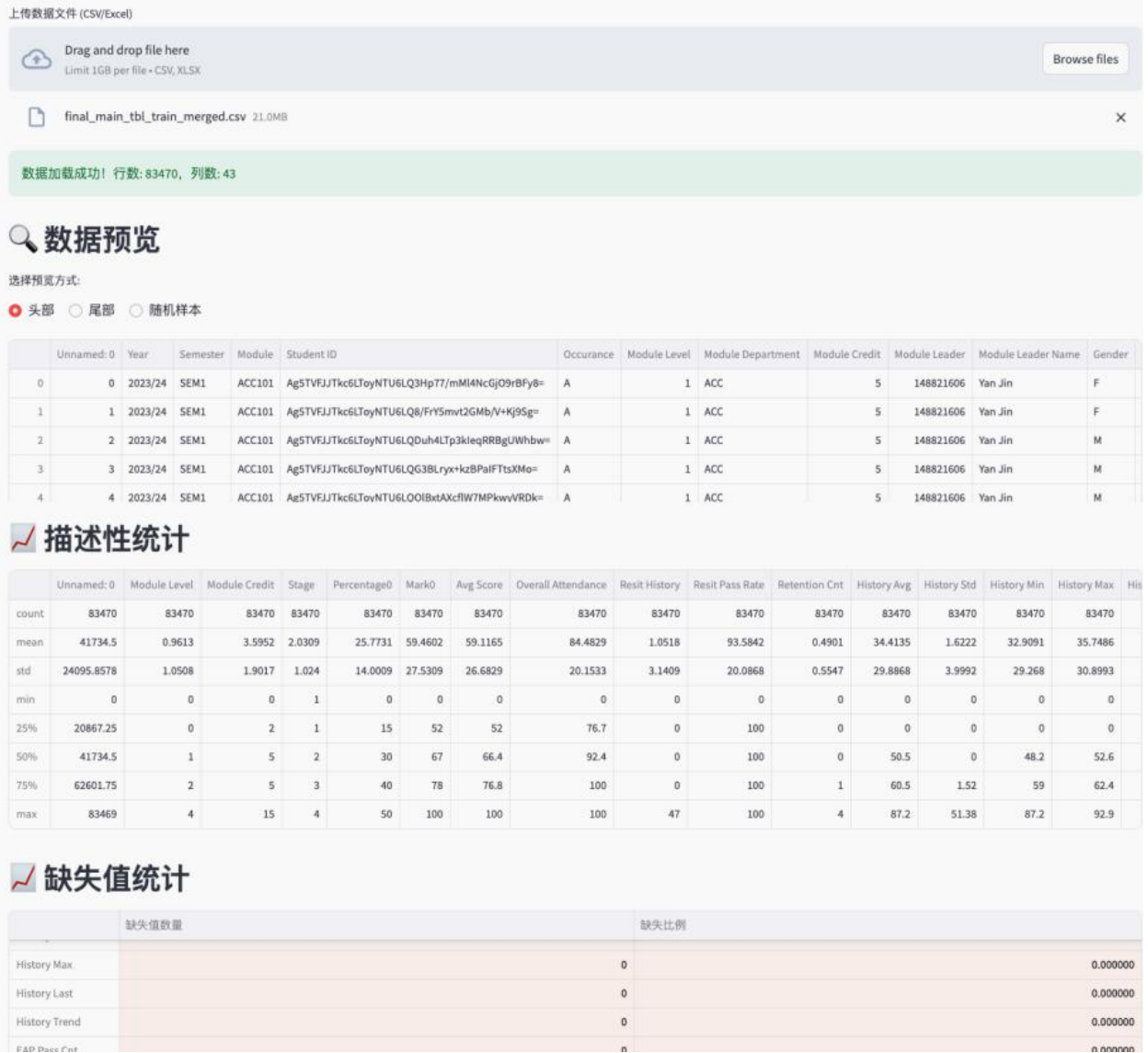


图 1 模型素具上传、预览和描述性统计



图2 相关性矩阵

```
False}. Best is trial 11 with value: 0.8686840104061164.
2025-11-13 00:09:57.419 | INFO | cf_train:optimize_hyperparameters:142 - best model value macf1 8.6868e-0
2025-11-13 00:09:57.437 | INFO | cf_model:create_model:140 - use model cat for regression? False ...
2025-11-13 00:12:23.220 | INFO | cf_train:optimize_hyperparameters:146 - save the model to ./model/cat#tr
in_risk_classification.csv#mdl.joblib
2025-11-13 00:12:23.244 | INFO | cf_train:test:206 - Loading model from ./model/cat#train_risk_classifica
ion.csv#mdl.joblib...
2025-11-13 00:12:23.646 | INFO | cf_train:test:234 - class dist in test set: #feat 49 #num 10504
2025-11-13 00:12:23.647 | INFO | cf_train:test:241 - Class distribution in test set:
2025-11-13 00:12:23.648 | INFO | cf_train:test:242 - PassModuleProb
0 7808
1 2696
Name: count, dtype: int64
2025-11-13 00:12:24.222 | INFO | cf_train:test:256 - Confusion Matrix:
2025-11-13 00:12:24.222 | INFO | cf_train:test:257 - [[7631 177]
 [ 336 2360]]
2025-11-13 00:12:24.276 | INFO | cf_train:test:266 - Accuracy: 0.9512
2025-11-13 00:12:24.276 | INFO | cf_train:test:267 - Micro F1-score: 0.9512
2025-11-13 00:12:24.284 | INFO | cf_train:test:268 - Macro F1-score: 0.9347
2025-11-13 00:12:24.386 | INFO | cf_train:test:274 - Classification Report:
precision recall f1-score support
0 0.96 0.98 0.97 7808
1 0.93 0.88 0.90 2696
accuracy 0.95 10504
macro avg 0.94 0.93 0.93 10504
weighted avg 0.95 0.95 0.95 10504
```

图3 模型性能 (recall&F1 分数)

(2) 数据安全保障

通过校内开发与数据本地化部署，彻底解决校外供应商的数据离域风险。

4. 可复制性与推广价值

(1) 模式创新

- **科研驱动管理**：以特殊科研项目整合跨部门资源，解决预算与协作壁垒，为高校同类项目提供参考；
- **跨学科协作**：融合教育学、心理学、计算机科学的研究框架，可推广至其他管理场景（如心理预警、就业预测）。

(2) 推广场景

适用于各类高校的学生风险管理系统建设，尤其适合需要整合多源数据、规避校外供应商风险的院校。