

Отчёт №3

第 3 次报告

1 Цель работы

Обработка кардиологического дата-сета для решения задач бинарной классификации

2 Метод

H2O AutoML, LightAutoML

3 Обсуждение

Данные были предварительно обработаны и очищены;

Фреймворки H2O AutoML и LightAutoML были выбраны для бинарного моделирования на основе результатов работы в программе-примере:

-Использовали фреймворк H2O AutoML, установили максимальное время обучения 60 секунд, автоматически выбрали подходящую модель из вариантов Gradient Boosting Machine (GBM), Random Forest, Deep Learning и Generalised Linear Model (GLM) и завершили настройку гиперпараметров; -Использовали предустановки TabularAutoML из LightAutoML, установили задачу на бинарную (Task('binary')), задали ограничение по таймауту и ресурсам CPU;

Вывели отчет о классификации, оценку F1, матрицу смещения, ROC-кривую, пороговую кривую F1 и другие графики.

1 目标

处理心脏病学数据集以解决二分类问题

2 方法

H2O AutoML 框架, LightAutoML 框架

3 操作

对数据进行了预处理和清洗

根据示例程序中的表现挑选了 H2O AutoML 框架和 LightAutoML 框架进行二分类建模:

· 使用了 H2OAutoML 框架, 设置了最大训练时间 60 秒, 自动从梯度提升机 (GBM)、随机森林、深度学习 (Deep Learning)、广义线性模型 (GLM) 等模型选择中选择合适的模型并完成了超参数调优

· 使用了 LightAutoML 中的 TabularAutoML 预设, 设置了任务为二分类 (Task('binary')), 设置了超时时长限制和 CPU 资源限制输出了分类报告、F1 分数, 生成了混淆矩阵、ROC 曲线、F1-阈值曲线等图表

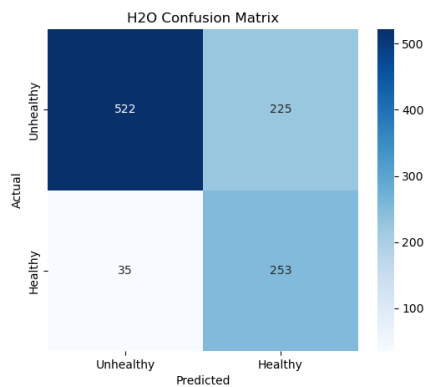


Figure 1: H2O Confusion Matrix

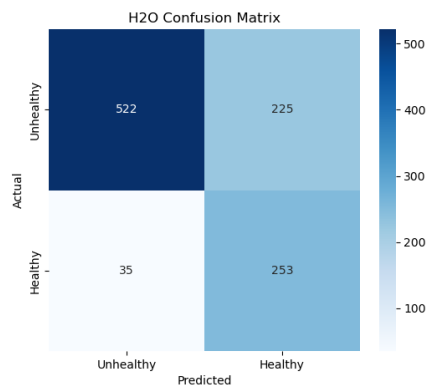


Figure 1: H2O 混淆矩阵

H2O Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Unhealthy | 0.94 | 0.70 | 0.80 | 747 |
| Healthy | 0.53 | 0.88 | 0.66 | 288 |
| accuracy | | | 0.75 | 1035 |
| macro avg | 0.73 | 0.79 | 0.73 | 1035 |
| weighted avg | 0.82 | 0.75 | 0.76 | 1035 |

Macro F1: 0.7306
Weighted F1: 0.7616
ROC AUC: 0.8457

Figure 2: H2O Report

H2O Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Unhealthy | 0.94 | 0.70 | 0.80 | 747 |
| Healthy | 0.53 | 0.88 | 0.66 | 288 |
| accuracy | | | 0.75 | 1035 |
| macro avg | 0.73 | 0.79 | 0.73 | 1035 |
| weighted avg | 0.82 | 0.75 | 0.76 | 1035 |

Macro F1: 0.7306
Weighted F1: 0.7616
ROC AUC: 0.8457

Figure 2: H2O 分类报告

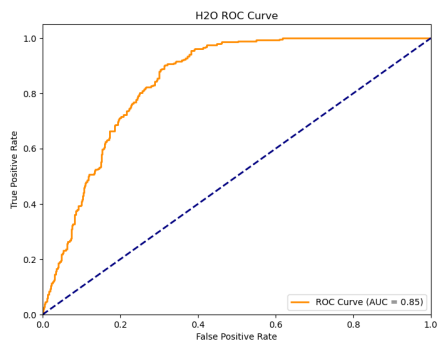


Figure 3: H2O Receiver Operating Characteristic Curve

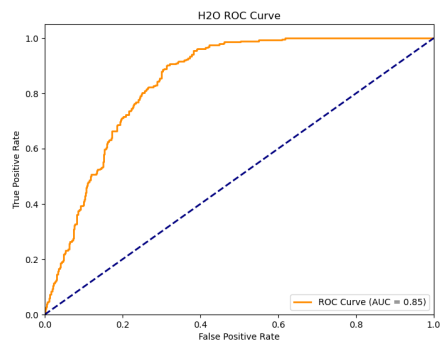


Figure 3: H2O ROC 曲线

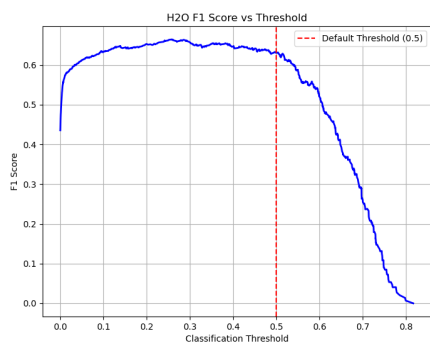


Figure 4: H2O F1 Score

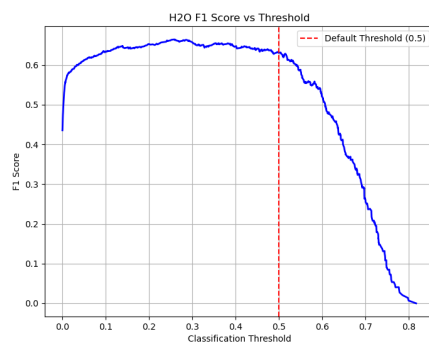


Figure 4: H2O F1 分数

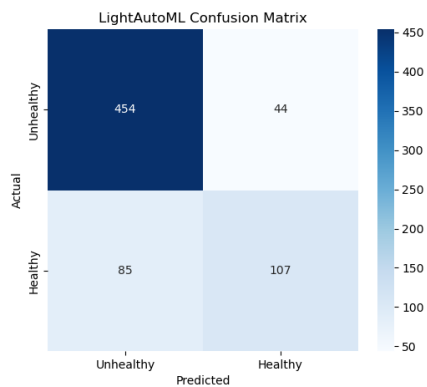


Figure 5: LightAutoML Confusion Matrix

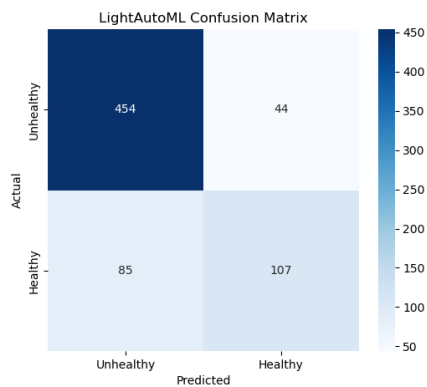


Figure 5: LightAutoML 混淆矩阵

LightAutoML Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Unhealthy | 0.84 | 0.91 | 0.88 | 498 |
| Healthy | 0.71 | 0.56 | 0.62 | 192 |
| accuracy | | | 0.81 | 690 |
| macro avg | 0.78 | 0.73 | 0.75 | 690 |
| weighted avg | 0.81 | 0.81 | 0.81 | 690 |

Macro F1: 0.7498
Weighted F1: 0.8056
ROC AUC: 0.8882

Figure 6: LightAutoML Report

LightAutoML Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Unhealthy | 0.84 | 0.91 | 0.88 | 498 |
| Healthy | 0.71 | 0.56 | 0.62 | 192 |
| accuracy | | | 0.81 | 690 |
| macro avg | 0.78 | 0.73 | 0.75 | 690 |
| weighted avg | 0.81 | 0.81 | 0.81 | 690 |

Macro F1: 0.7498
Weighted F1: 0.8056
ROC AUC: 0.8882

Figure 6: LightAutoML 分类报告

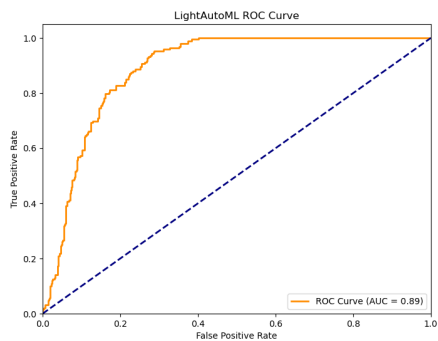


Figure 7: LightAutoML Receiver Operating Characteristic Curve

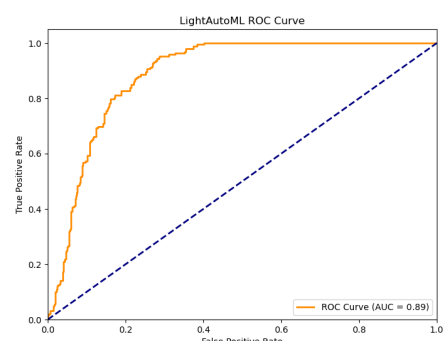


Figure 7: LightAutoML ROC 曲线

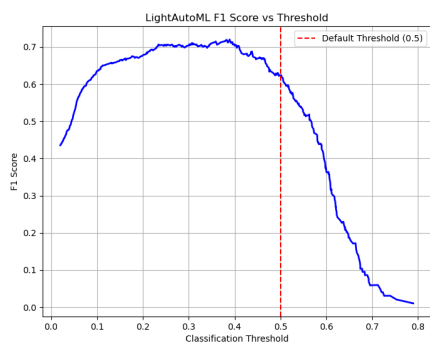


Figure 8: LightAutoML F1 Score

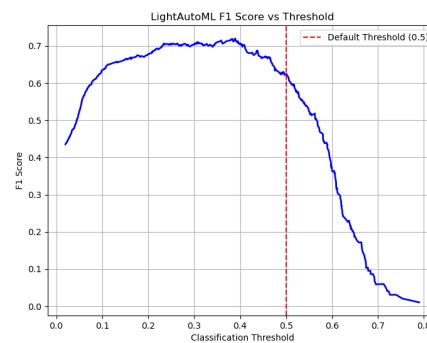


Figure 8: LightAutoML F1 分数

4 Ссылки на литературу

References

- [1] *Набор данных для лабораторных работы и исследований* сайт. 2025. URL: <https://github.com/AI-is-out-there/data2lab> (Дата обращения: 16.03.2025)
- [2] *ML ECG classification:* сайт. 2025. URL: <https://github.com/TAUforPython/BioMedAI/blob/main/ML%20ECG%20classification.ipynb> (Дата обращения: 06.04.2025)
- [3] *H2O AutoML: Automatic Machine Learning:* сайт. 2025. URL: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html> (Дата обращения: 06.04.2025)
- [4] *LightAutoML documentation:* сайт. 2025. URL: <https://lightautoml.readthedocs.io/en/v.0.4.0/> (Дата обращения: 06.04.2025)
- [5] *Summary of common model evaluation metrics for machine learning models:* сайт. 2025. URL: https://blog.csdn.net/weixin_42521211/article/details/144224113 (Дата обращения: 06.04.2025)

4 参考文献

References

- [1] AI-is-out-there. Набор данных для лабораторных работы и исследований[EB/OL]. (2025-03-16)[2025-04-06]. <https://github.com/AI-is-out-there/data2lab>
- [2] TAUforPython. ML ECG classification[EB/OL]. (2025-04-06)[2025-04-06]. <https://github.com/TAUforPython/BioMedAI/blob/main/ML%20ECG%20classification.ipynb>
- [3] H2O.ai. H2O AutoML: Automatic Machine Learning[EB/OL]. (2025-04-06)[2025-04-06]. <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>
- [4] LightAutoML Team. LightAutoML documentation[EB/OL]. (2025-04-06)[2025-04-06]. <https://lightautoml.readthedocs.io/en/v.0.4.0/>
- [5] CSDN. Summary of common model evaluation metrics for machine learning models[EB/OL]. (2025-04-06)[2025-04-06]. https://blog.csdn.net/weixin_42521211/article/details/144224113