

Отчёт 3

ИУ1И-41М Цен Лэшань

28 апреля 2025г.

1 Цель работы

Обработайте наборы кардиологических данных, чтобы решить проблему двух классификаций.

Статус работоспособности классифицируется с помощью методов автоматизированного машинного обучения (AutoML). А также оценить производительность модели и продемонстрировать выбор наилучшего фреймворка.

1 目标

处理心脏病学数据集以解决二分类问题。

通过自动化机器学习（AutoML）方法对健康状况（Healthy Status）进行分类。并评估模型的性能，论证最佳框架的选择。

2 Метод

Датасет

Загрузите 5000 наборов данных ЭКГ с GitHub. Будут выбраны необходимые столбцы и первые 5000 записей.

2 方法

数据集

从 GitHub 加载 5000 行 ECG 数据集。选择了必需的列和前 5000 条记录。

Обучите и оцените модель

AutoML сокращает ручное вмешательство за счет автоматического выбора функций, модели и настройки гиперпараметров. Используйте H2OAutoML для автоматического обучения нескольких моделей. Используйте таблицу лидеров,

训练并评估模型

AutoML 通过自动化特征选择、模型选择和超参数调优，减少人工干预。使用 H2OAutoML 自动训练多个模型。通过 leaderboard 对比模型性能，选择最优模型。

чтобы оценить производительность модели и выбрать оптимальную модель.

Результаты обучения

Используйте автоматическое обучение H2O, установите максимальное количество моделей равным 10.

Получите оптимальную модель. Рассчитайте матрицу путаницы, чтобы отобразить эффект классификации.

Тренировочные результаты

Используйте H2OAutoML для тренировки, установите максимальное количество моделей равным 10.

Получите оптимальную модель. Рассчитайте матрицу путаницы (Confusion Matrix), чтобы отобразить эффект классификации.

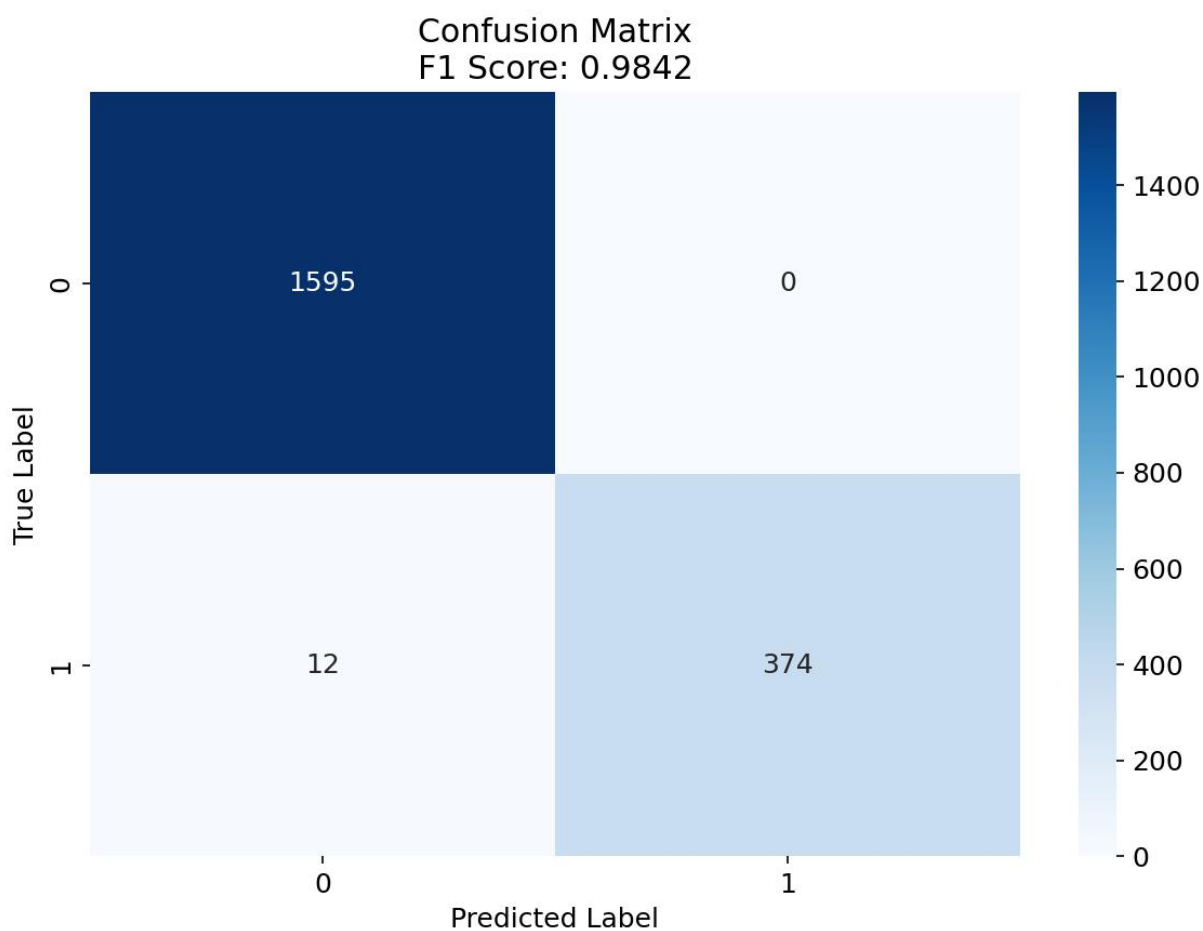


Рис.1 Матрица путаницы 混淆矩阵

```

=====
最佳模型: GBM_2_AutoML_1_20250514_83010
模型参数摘要:
dict_keys(['model_id', 'training_frame', 'validation_frame', 'nfolds', 'keep_cross_validation_models', 'keep_cross_validation_predictions',
=====

模型排行榜:
model_id          auc      logloss    aucpr    mean_per_class_error    rmse      mse
GBM_2_AutoML_1_20250514_83010  0.999917  0.00459405  0.999741    0.00123381  0.0249393  0.00062197
GBM_3_AutoML_1_20250514_83010  0.999872  0.00452649  0.999706    0.00108165  0.0249439  0.000622197
GLM_1_AutoML_1_20250514_83010  0.999768  0.0382176   0.999195    0.00732883  0.0797547  0.00636082
GBM_1_AutoML_1_20250514_83010  0.999763  0.00475665  0.999442    0.00154226  0.0251315  0.000631594
DRF_1_AutoML_1_20250514_83010  0.999569  0.0122139   0.999549    0.000851351  0.0364555  0.001329
[5 rows x 7 columns]

```

Рис.2 Лучшая модель 最佳模型

3 Заключение

В работе использовался AutoML для обработки наборов кардиологических данных и был создан бинарный классификатор для определения наилучшей модели на основе оценки F1. AutoML хорошо справляется с автоматизацией выбора модели и настройки параметров и подходит для обработки задач классификации данных ЭКГ.

3 结论

工作使用了 AutoML 处理心脏病学数据集，并构建二元分类器，基于 F1 分数确定了最佳模型。AutoML 在自动化模型选择和调参上表现优异，适合处理 ECG 数据的分类任务。

References

<https://github.com/TAUforPython/BioMedAI/blob/main/ML%20ECG%20classification.ipynb>