

Отчёт 报告

Исса Мохаммад 伊萨·穆罕默德

13 мая 2025 г. 2025 年 5 月 13 日

1 Цель работы

Целью данной работы является автоматическая классификация ЭКГ-записей по параметру здоровья пациента (`Healthy_Status`) с использованием современных AutoML-фреймворков. Анализ выполняется на базе реальных медицинских данных, включающих временные и угловые параметры электрокардиограммы.

2 Метод

Выбор фреймворка. Среди популярных AutoML-решений были рассмотрены FLAML, H2O AutoML и LightAutoML. Для решения задачи бинарной классификации был выбран фреймворк FLAML, поскольку он продемонстрировал наилучшие результаты при ограниченном времени обучения (60 секунд), что особенно важно в условиях ограниченных вычислительных ресурсов. FLAML автоматически подбирает оптимальную модель и её гиперпараметры, ориентируясь на метрику F1-score. В рамках эксперимента также были построены ROC-кривая и матрица ошибок для оценки качества классификации.

3 Обсуждение

Исходный набор данных содержал более 30 признаков. После этапов очистки и отбора были оставлены ключевые переменные: `Count_subj`, `rr_interval`, `p_end`, `qrs_onset`, `qrs_end`, `p_axis`, `qrs_axis`, `t_axis`.

В результате обучения классификатор продемонстрировал следующие показатели:

- Точность (Accuracy): 85%

1 目标

本研究旨在利用现代的 AutoML 框架，根据患者的 `Healthy_Status` 参数对心电图 (ECG) 记录进行自动分类。分析基于真实的医疗数据，包含心电图的时间和角度参数。

2 方法

框架选择。本研究对主流的 AutoML 框架进行了评估，包括 FLAML、H2O AutoML 和 LightAutoML。最终选择了 FLAML 来解决二分类任务，因为它在仅限 60 秒训练时间的条件下表现最佳，尤其适用于计算资源有限的场景。FLAML 能够根据 F1 分数自动选择最优模型和超参数。在实验中还绘制了 ROC 曲线和混淆矩阵，以评估分类器的性能。

3 讨论

原始数据集中包含了 30 多个特征变量。经过数据清洗和筛选后，保留了以下关键特征: `Count_subj`、`rr_interval`、`p_end`、`qrs_onset`、`qrs_end`、`p_axis`、`qrs_axis`、`t_axis`。训练得到的分类模型具有以下性能指标:

- 准确率 (Accuracy): 85%
- F1 分数 (类别 1): 0.64
- AUC 值: 0.92

- F1-мера для класса 1: 0.64
- Площадь под ROC-кривой (AUC): 0.92

Полученные результаты свидетельствуют о том, что модель способна эффективно различать два класса, несмотря на частичное перекрытие распределений признаков.

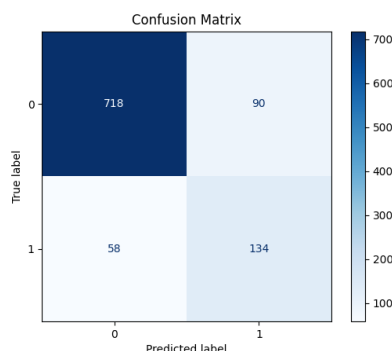


Figure 1: Матрица ошибок (Confusion Matrix)

Матрица ошибок отображает распределение предсказанных и фактических меток классов модели. Высокая концентрация значений на диагонали указывает на хорошую точность классификации.

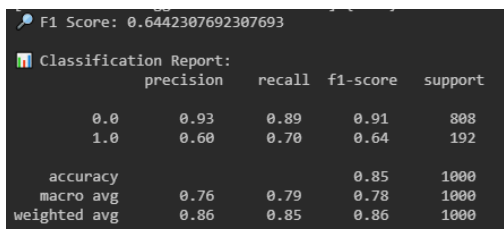


Figure 2: Отчёт классификации

Классификационный отчёт предоставляет значения точности, полноты и F1-меры для каждого класса. Он позволяет оценить сбалансированность модели по метрикам.

Это表明尽管两个类别在数据上存在重叠，模型仍能成功区分它们。

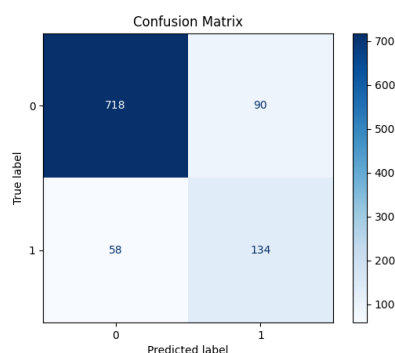


Figure 1: 混淆矩阵

混淆矩阵展示了模型预测结果与实际类别之间的对应关系。对角线上的高密度表示分类器具有良好的准确性。

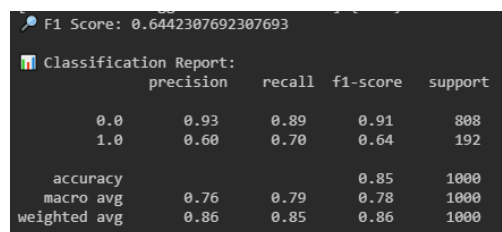


Figure 2: 分类报告

分类报告列出了每个类别的精确率、召回率和 F1 分数，用于评估模型的性能与平衡性。

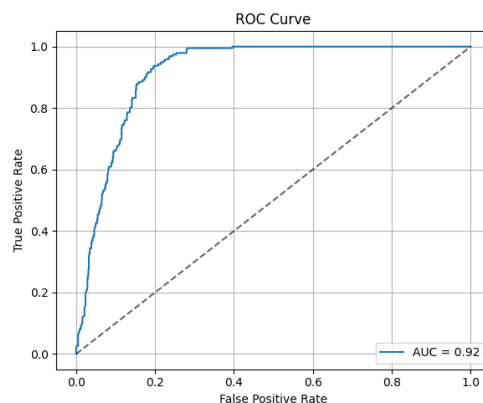


Figure 3: ROC 曲线

ROC 曲线展示了模型在不同阈值下的真阳性率与假阳性率之间的权衡。AUC 值为 0.92 表明模型区分类别的能力很强。

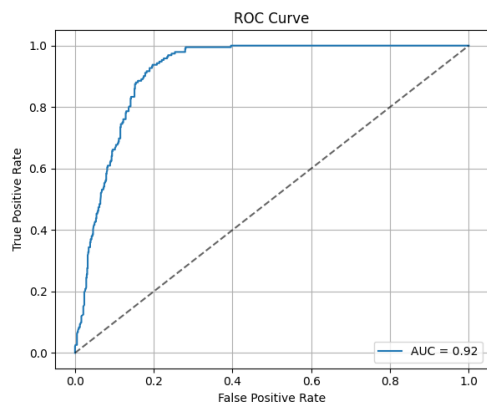


Figure 3: ROC-кривая

ROC-кривая иллюстрирует соотношение между истинно положительными и ложноположительными результатами. Площадь под кривой (AUC) равная 0.92 говорит о высокой способности модели различать классы.

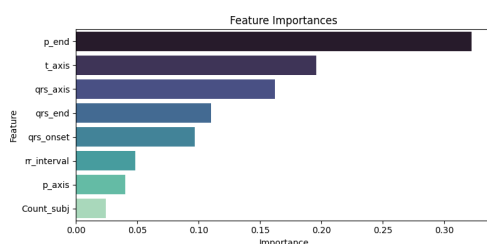


Figure 4: Важность признаков

График важности признаков демонстрирует вклад каждого входного параметра в предсказание модели. Наиболее значимые признаки оказывают наибольшее влияние на результат классификации.

4 Выводы

Проведённый анализ подтвердил, что AutoML-фреймворк FLAML способен эффективно решать задачу классификации ЭКГ-записей с высокой точностью (AUC 0.92). Использование автоматизированных подходов сокращает время настройки модели и повышает воспроизводимость результатов.

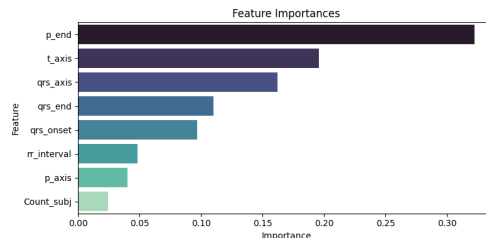


Figure 4: 特征重要性

特征重要性图反映了各输入变量对模型预测的贡献程度。得分较高的特征对最终分类结果影响最大。

4 结论

实验结果表明，FLAML 框架可以有效地完成心电图分类任务，AUC 值高达 0.92，表明模型具有很强的判别能力。自动机器学习显著简化了建模流程，提高了结果的稳定性与效率。

5 Ссылки на литературу

- [1] Набор данных для лабораторных работы и исследований. 2025. URL: <https://github.com/AI-is-out-there/data2lab> (Дата обращения: 16.03.2025)
- [2] *FLAML: A Fast Library for AutoML*. 2025. URL: <https://github.com/microsoft/FLAML> (Дата обращения: 06.05.2025)

5 参考文献

- [1] AI-is-out-there. 心脏数据集 [EB/OL]. (2025-03-16) [2025-05-06]. <https://github.com/AI-is-out-there/data2lab>
- [2] Microsoft. FLAML 自动机器学习库 [EB/OL]. (2025-05-06) [2025-05-06]. <https://github.com/microsoft/FLAML>