### ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук Образовательная программа «Машинное обучение и высоконагруженные системы»

#### ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Исследовательский проект на тему:

Сравнительное исследование AutoML фреймворков на табличных данных

#### Выполнил студент:

группы ММОВС22 Плющ Евгений Денисович

#### Принял руководитель ВКР:

Кантонистова Елена Олеговна к ф.-м.н, доцент Факультет компьютерных наук НИУ ВШЭ

# Содержание

A	нот	ация		3		
1	Вве	едение		5		
2	тературы	6				
	2.1	Auto-	sklearn	6		
	2.2	Oboe		6		
	2.3	Auto	Gluon	6		
	2.4	Light	AutoML	7		
	2.5	Проч	ие решения	7		
3	Исс	ледов	ание	ę		
	3.1	Описа	ание проведенных экспериментов	(		
		3.1.1	Данные	Ć		
		3.1.2	План экспериментов	10		
		3.1.3	Бейзлайн (Baseline)	11		
		3.1.4	Стек	11		
		3.1.5	Инфраструктура для экспериментов	11		
	3.2	Резул	ьтаты	12		
		3.2.1	Secondary Mushroom	12		
		3.2.2	Apple Quality	13		
		3.2.3	Credit Card Fraud Detection Dataset 2023	13		
		3.2.4	Census Income	14		
		3.2.5	CDC Diabetes Health Indicators	14		
		3.2.6	Sepsis Survival Minimal Clinical Records	15		
	3.3	Выво,	ды	15		
4	Зак	лючеі	ние	17		
Cı	писо	к лите	ературы	18		
$\Pi_1$	рило	жения	a.	<b>2</b> 1		
	А Диаграммы размаха для проведенных экспериментов					

## Аннотация

В последние годы в машинном обучении активно развивается подход автоматизированного машинного обучения (AutoML). Для табличных данных этот подход приобрел наибольшую популярность, поскольку однородность структуры позволяет упростить автоматизацию процессов. В работе проведено сравнение нескольких AutoML фреймворков в задаче бинарной классификации на табличных данных в условиях ограниченных ресурсов.

# Ключевые слова

AutoML, автоматизированное машинное обучение, классификация, табличные данные

## 1 Введение

В последние годы машинное обучение применяется все чаще в разных сферах. Несмотря на популярность задач компьютерного зрения, генерации и классификации текстов [1], наиболее часто встречающаяся форма данных - табличная [2]. Структура таких данных однородна, что упрощает автоматизацию процессов, поэтому появление автоматизированных подходов к поиску подходящих моделей было вопросом времени [3] - и значительный рост числа публикаций с 2018 года лишь подтверждает эту мысль.

Введем определение **автоматизированного машинного обучения (AutoML)** на основе источника [1]. *AutoML* - это пайплайн машинного обучения, автоматизирующий процессы предобработки, создания признаков, генерации модели (этап, состоящий из выбора моделей, поиска гиперпараметров и оптимизации), а также валидации модели.

На момент написания работы наибольшее распространение имеют несколько активно развивающихся фреймворков, которые рассмотрены в работе, но в сообществе по-прежнему нет однозначного отношения к ним [1]. На данный момент AutoML не всегда способны превзойти решения от экспертов, поскольку зачастую недостаточно только эмпирических правил и формул для подбора лучшего решения [4]. Вопрос, сможет ли AutoML со временем полностью вытеснить экспертные решения, является дискуссионным и останется за рамками данной работы.

Целью работы является сравнение AutoML фреймворков между собой и с "человеческим" решением в условиях ограниченности ресурсов.

## 2 Обзор литературы

В этом разделе приведен краткий обзор работ, посвященных AutoML фреймворкам.

#### 2.1 Auto-sklearn

Автоматизированное машинное обучение начало свое развитие около 2014 года [3] с зарождением фреймворка Auto-Sklearn [5] Особенностью стал опыт мета-обучения (рисунок 2.1): Auto-sklearn содержит информацию об обучении на 15 датасетах, которую использует для подбора пайплайна

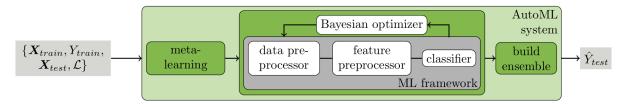


Рис. 2.1: Общая схема пайплайнов Auto-sklearn [5]

#### 2.2 Oboe

Следующим знаковым этапом стало появление фреймворка Oboe [6] в 2018 году и его улучшения TensorOboe [7] в 2019. Особенностью Oboe является использование матричных разложений (рисунок 2.2) для идеи из Auto-sklearn (мета-обучения). Проблема Oboe состоит в реализации, которая на момент написания работы непригодна к использованию.

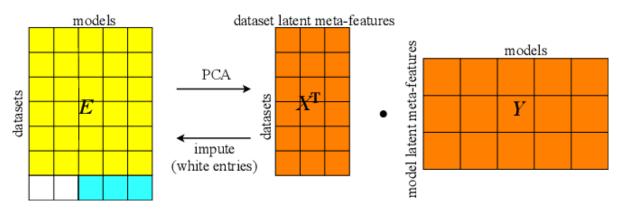


Рис. 2.2: Схема из оригинальной статьи [6]

#### 2.3 AutoGluon

Еще один значимый этап - появление AutoGluon [8] - фреймворка, выпущенного в ореп-source компанией Amazon. Особенностью AutoGluon стало использование многоэтапных стекингов (рисунок 2.3) лучших моделей для табличных данных - бустингов и нейросетей [2].

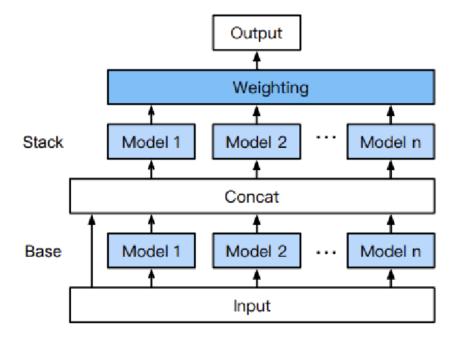


Figure 2. AutoGluon's multi-layer stacking strategy, shown here using two stacking layers and n types of base learners.

Рис. 2.3: Схема многоэтапного стекинга из статьи [8]

## 2.4 LightAutoML

Одним из самых популярных AutoML фреймворков является LightAutoML [9] (LAMA), появившийся в 2021 году. Имеет обширную кодовую базу с примерами промышленного применения и поддержку <sup>1</sup> с обучающими материалами сообщества <sup>2</sup>. LightAutoML собрал в себе особенности предшественников, но особенностью стал отказ от применения нейросетей. Даже несмотря на это, по словам авторов статьи, LightAutoML в соревновательном DS показывает себя лучше, чем 90% экспертов<sup>3</sup>.

## 2.5 Прочие решения

Также стоит упомянуть несколько решений, не участвующих в сравнении в этой работе

#### • H2O 4

https://developers.sber.ru/portal/products/lightautoml

<sup>2</sup>https://ods.ai/tracks/automl-course-part1

<sup>&</sup>lt;sup>3</sup>https://github.com/sberbank-ai-lab/Paper-Appendix

<sup>4</sup>http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html

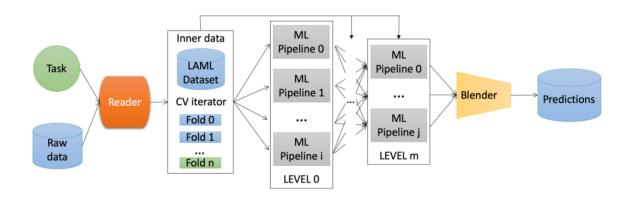


Рис. 2.4: Схема работы LightAutoML [9]

H2O - экосистема ML-решений с поддержкой нескольких языков, включающая в себя AutoML-компонент. Несмотря на масштабируемость, имеет проблемы с производительностью в условиях ограниченных ресурсов.

#### • AutoPytorch [10]

AutoML фреймворк для автоматического подбора сугубо нейросетевых архитектур от создателей Auto-sklearn. Требует лимит по времени значительно больший, чем ограничения в данной работе.

## 3 Исследование

В этом разделе описаны методика проведения экспериментов и результаты, а также выводы из них.

### 3.1 Описание проведенных экспериментов

#### 3.1.1 Данные

Для исследования были выбраны известные датасеты с задачей бинарной классификации с разными размерностями (количествами признаков и числом объектов).

- Sepsis Survival Minimal Clinical Records [11]
- Census Income [12]
- CDC Diabetes Health Indicators [13]
- Secondary Mushroom [14]
- Credit Card Fraud Detection Dataset 2023 <sup>5</sup>
- Apple Quality <sup>6</sup>

Подробные характеристики датасетов представлены в таблице 3.1. В качестве метрики для всех датасетов используется ROC-AUC

Таблица 3.1: Список использованных датасетов

Название	Число объек-	Число призна-	Особенности	
	тов	ков		
Sepsis Survival Minimal	110k	3	Мало признаков, один	
Clinical Records			категориальный	
Census Income	48k	14	8 категориальных при-	
			знаков	
CDC Diabetes Health	253k	21	Только целочисленные	
Indicators			(категориальные и ран-	
			жирующие) признаки	

 $<sup>^5 \</sup>rm https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023, дата обр. <math display="inline">16.05.2024$ 

<sup>&</sup>lt;sup>6</sup>https://www.kaggle.com/datasets/nelgiriyewithana/apple-quality/data, дата обр. 16.05.2024

Таблица 3.1: Список использованных датасетов

Название	Число объек-	Число призна-	Особенности
	тов	ков	
Secondary Mushroom	61k	20	Синтетический датасет
			на основе реального
			Primary Mushroom [15]
			17 категориальных
			признаков
Credit Card Fraud	568k	29	29 числовых признаков,
Detection Dataset 2023			из них 28 неинтерпрети-
			руемых
Apple Quality	4k	7	7 нормированных дей-
			ствительных признаков

#### 3.1.2 План экспериментов

Под каждый фреймворк создается своя среда (conda [19] env) в docker-контейнере [16] на одном и том же хосте с характеристиками:

- Процессор AMD Ryzen 9 8945h с тактовой частотой 4GHz и 16 сри-ядрами
- Оперативная память **32Gb** DDR5

#### Условия для запусков (ограничения контейнеров)

- Ограничение по памяти 8Gb
- Ограничение по сри 4 ядра
- Отключен доступ к gpu

#### Условия обучения

- Используется только "out of the box"применение (без процедур предобработок)
- Отсчет времени стартует с момента начала работы фреймворка (после инициализаций)
- Все рандомизации фиксируются (fixed seed) для воспроизводимости

• Обучение происходит на 80% данных, валидация на 20% отложенных данных

Общую схему одного эксперимента описывает алгоритм 1

#### Алгоритм 1 Алгоритм эксперимента

```
Initialize dataset, AutoML, maxtime N \leftarrow 10

for i = 0 to N do

X_{eval}, y_{eval}, X_{train}, y_{train} \leftarrow X, y split with fixed seed time \leftarrow 0

while time < maxtime do

AutoML training process
end while
metric \leftarrow AutoML(X_{eval}, y_{eval})
append metric to results
end for return results
```

Многократные запуски позволяют снизить влияние сторонних факторов (подходящего разбиения датасета, удачной инициализации). Получаемые значения агрегируются для сравнения (используется среднее арифметическое с учетом стандартного отклонения).

#### 3.1.3 Бейзлайн (Baseline)

В качестве бейзлайна в виде "экспертного решения от человека" было принято решение использовать Catboost [17] с перебором гиперпараметров с помощью Optuna [18] при аналогичных лимитах

#### 3.1.4 Стек

Использованный стек технологий и библиотеки:

- Окружения: Python 3.8-3.11 (в зависимости от среды), Docker [16], Conda [19]
- Мониторинг и сбор метрик: Weights&Biases [20]
- Обучение: Scikit-learn [21], Catboost [17], Pandas [22], Numpy [23], Optuna [18]
- Визуализация: Matplotlib [24], Seaborn [25]

#### 3.1.5 Инфраструктура для экспериментов

Для экспериментов была реализована инфраструктура на Python. Ограничения по ресурсам достигаются за счет использования докер-контейнеров. Схематичное изображение

архитектуры представлено на рисунке 3.1. Весь написанный код находится в открытом доступе в репозитории<sup>7</sup>

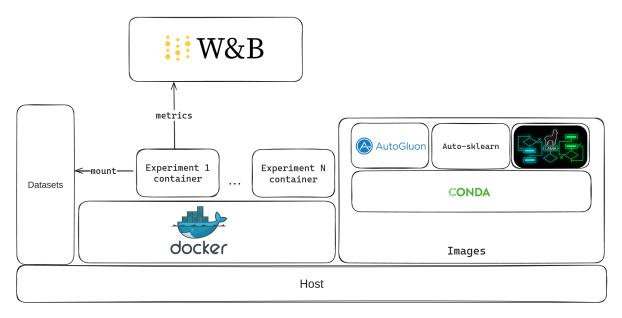


Рис. 3.1: Инфраструктура для экспериментов

#### 3.2 Результаты

В данной секции приведены результаты экспериментов по датасетам с интерпретацией. Все диаграммы разброса для каждого эксперимента можно найти в Приложении A

#### 3.2.1 Secondary Mushroom

Датасет для проверки работоспособности фреймворков. По таблице 3.2 видно, что Auto-sklearn не справляется с данной задачей в сжатые сроки, что в принципе является проблемой фреймворка. Далее он исключен из сравнений там, где не успевал подобрать хотя бы одно решение.

Таблица 3.2: Результаты для Secondary Mushroom

Ограничение по	Auto-sklearn	AutoGluon	Catboost&Optuna	LightAutoML
времени				
30s	-	1	1	1

<sup>7</sup>https://github.com/Zhekuson/AutoML-thesis

#### 3.2.2 Apple Quality

Малый датасет, в котором признаки - нормированные действительные числа. В таблице 3.3 видно, что имеет качество выше AutoGluon, но решение от LAMA стабильнее, поскольку имеет меньший разброс (рисунок 3.2). Также наблюдается более быстрый выход на плато по качеству у LAMA, чем у бейзлайна и AutoGluon.

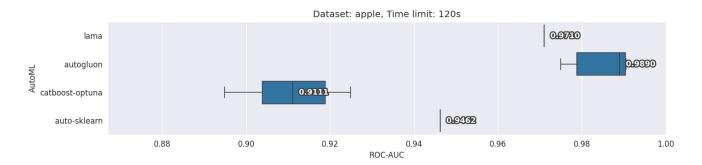


Рис. 3.2: Диаграммы размаха качества, датасет Apple Quality, ограничение 120 секунд

Ограничение по	Auto-sklearn	AutoGluon	Catboost&Optuna	LightAutoML
времени				
30s	0.9225±0.0000	$0.9762 \pm 0.0041$	$0.8969 \pm 0.0141$	$0.9690 \pm 0.0000$
60s	0.9412±0.0000	$0.9725 \pm 0.0057$	$0.9056 \pm 0.0086$	$0.9709 \pm 0.0001$
90s	$0.9416 \pm 0.0012$	$0.9710 \pm 0.0022$	$0.9109 \pm 0.0065$	$0.9710 \pm 0.0000$
120s	0.9462±0.0000	$0.9849 \pm 0.0066$	$0.9109 \pm 0.0101$	$0.9710 \pm 0.0000$
150s	$0.9448 \pm 0.0020$	$0.9717 \pm 0.0017$	$0.9108 \pm 0.0036$	$0.9710 \pm 0.0000$

Таблица 3.3: Результаты для Apple Quality

#### 3.2.3 Credit Card Fraud Detection Dataset 2023

В этом датасете 29 числовых признаков с действительными значениями. Здесь нахождение идеального решения LAMA происходит при минимальном лимите времени (таблица 3.4). AutoGluon, напротив, не может по качеству преодолеть бейзлайн.

Таблица 3.4: Результаты для Credit Card Fraud Detection Dataset 2023

Ограничение по	AutoGluon	Catboost&Optuna	LightAutoML
времени			
30s	$0.9677 \pm 0.0000$	$0.9983 \pm 0.0021$	1.0000±0.0000
60s	$0.9689 \pm 0.0013$	$0.9990 \pm 0.0010$	1.0000±0.0000

Таблица 3.4: Результаты для Credit Card Fraud Detection Dataset 2023

Ограничение по	AutoGluon	Catboost&Optuna	LightAutoML
времени			
90s	0.9787±0.0116	$0.9994 \pm 0.0011$	1.0000±0.0000

#### 3.2.4 Census Income

На этом датасете LAMA приходит к своему лучшему решению при минимальном лимите времени (таблица 3.5). Также видно, что AutoGluon не может преодолеть бейзлайн по качеству. Auto-sklearn успевает найти решение только за 150 секунд, и оно сильно проигрывает по качеству остальным

Таблица 3.5: Результаты для Census Income

Ограничение по	Auto-sklearn	AutoGluon	Catboost&Optuna	LightAutoML
времени				
30s	-	$0.9073 \pm 0.0033$	$0.9198 \pm 0.0021$	$0.9306 \pm 0.0000$
60s	-	$0.9126 \pm 0.0066$	$0.9205 \pm 0.0015$	$0.9307 \pm 0.0000$
90s	-	$0.9129 \pm 0.0024$	0.9201±0.0025	$0.9308 \pm 0.0001$
120s	-	$0.9190 \pm 0.0044$	$0.9208 \pm 0.0013$	0.9307±0.0000
150s	0.7954±0.0016	$0.9169 \pm 0.0016$	$0.9219 \pm 0.0009$	0.9307±0.0000

#### 3.2.5 CDC Diabetes Health Indicators

На этом датасете результаты двух фреймворков и бейзлайна практически совпадают (таблица 3.6).

Таблица 3.6: Результаты для CDC Diabetes Health Indicators

Ограничение по	AutoGluon	Catboost&Optuna	LightAutoML
времени			
30s	$0.8162 \pm 0.0036$	$0.8225 \pm 0.0115$	$0.8291 \pm 0.0001$
60s	$0.8208 \pm 0.0045$	$0.8275 \pm 0.0047$	$0.8291 \pm 0.0000$
90s	$0.8206 \pm 0.0011$	$0.8284 \pm 0.0020$	$0.8291 \pm 0.0000$
120s	$0.8245 \pm 0.0024$	$0.8288 \pm 0.0018$	0.8292±0.0000
150s	$0.8249 \pm 0.0010$	$0.8301 \pm 0.0018$	0.8292±0.0000

#### 3.2.6 Sepsis Survival Minimal Clinical Records

Датасет с малым числом признаков. По таблице 3.7 наблюдается практически совпадающие метрики LAMA и AutoGluon, а также высокий разброс у бейзлайна (рисунок 3.3).

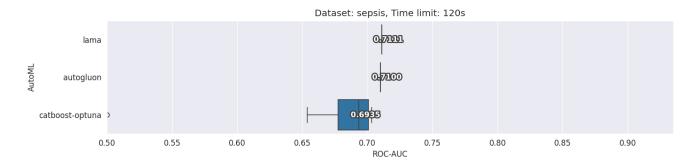


Рис. 3.3: Диаграммы размаха качества, датасет Sepsis Survival Minimal Clinical Records, ограничение 120 секунд

AutoGluon Ограничение Catboost&Optuna LightAutoML ПО времени  $0.7100 \pm 0.0004$  $0.7106 \pm 0.0000$ 30s $0.6795 \pm 0.0689$ 60s $0.7109\pm0.0006$  $0.7010\pm0.0050$  $0.7111 \pm 0.0000$ 90s $0.7101\pm0.0004$  $0.6724 \pm 0.0625$  $0.7111 \pm 0.0000$ 120s $0.7100\pm0.0001$  $0.6935 \pm 0.0619$  $0.7111 \pm 0.0000$ 

Таблица 3.7: Результаты для Sepsis Survival Minimal Clinical Records

### 3.3 Выводы

По итогу проведенных экспериментов можно сделать несколько выводов:

- AutoML фреймворки все еще недостаточно хороши по сравнению с экспертными решениями. Не всегда стабильно удается победить простые бейзлайны по качеству, хотя прогресс в этом направлении имеется. Некоторые фреймфорки работают очень медленно (AutoPytorch, H2O) и при этом выдают низкое качество (Auto-sklearn), некоторые плохо поддерживаются и не готовы к промышленному применению (Oboe).
- AutoML фреймворки стабильнее, чем экспертные решения. Экспертные решения гораздо богаче по наполнению и идеям, но даже простой бейзлайн из этой работы показывал высокий разброс по качеству. В выборе между AutoML и человеческими решениями этот фактор стоит учитывать.

- В условиях ограниченных ресурсов для получения решения достаточного качества стоит выбирать LightAutoML
- AutoGluon может выдать качество лучше, чем LightAutoML, но при этом имеет больший разброс, а LightAutoML быстро выходит на плато и придерживается выбранного ансамбля моделей даже с ростом лимита по времени, поэтому AutoGluon больше подходит для экспериментов и перебора вариантов решения.

### 4 Заключение

В данной работе происходило сравнение AutoML фреймворков в условиях ограниченных ресурсов. Была проведены эксперименты, построена инфраструктура для их проведения. Ключевые выводы из работы:

- AutoML фреймворк LightAutoML чаще всего показывает самое стабильное решение в условиях ограниченных ресурсов
- AutoML фреймворк AutoGluon тоже может показывать высокий результат
- При этом простой "экспертный" бейзлайн Catboost+Optuna показывает менее стабильный по сравнению с AutoML результат, иногда побеждая AutoML

## Список литературы

- Xin He, Kaiyong Zhao u Xiaowen Chu. "AutoML: A survey of the state-of-the-art". B: Knowledge-Based Systems 212 (2021), c. 106622. ISSN: 0950-7051. DOI: https://doi.org/ 10.1016/j.knosys.2020.106622. URL: https://www.sciencedirect.com/science/ article/pii/S0950705120307516.
- 2. Ravid Shwartz-Ziv и Amitai Armon. "Tabular data: Deep learning is not all you need". B: Information Fusion 81 (2022), c. 84—90.
- 3. Rafael Barbudo, Sebastián Ventura и José Raúl Romero. "Eight years of AutoML: categorisation, review and trends". B: *Knowledge and Information Systems* 65.12 (2023), c. 5097—5149.
- 4. Salisu Mamman Abdulrahman, Pavel Brazdil, Jan N van Rijn и Joaquin Vanschoren. "Speeding up algorithm selection using average ranking and active testing by introducing runtime". B: *Machine learning* 107 (2018), c. 79—108.
- 5. Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum и Frank Hutter. "Efficient and Robust Automated Machine Learning". B: Advances in Neural Information Processing Systems 28 (2015). 2015, с. 2962—2970.
- 6. Chengrun Yang, Yuji Akimoto, Dae Won Kim и Madeleine Udell. "OBOE: Collaborative filtering for AutoML model selection". B: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019, c. 1173—1183.
- 7. Chengrun Yang, Jicong Fan, Ziyang Wu 
  Madeleine Udell. "Automl pipeline selection: Efficiently navigating the combinatorial space". B: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020, c. 1446—1456.
- 8. Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li и Alexander Smola. "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data". B: arXiv preprint arXiv:2003.06505 (2020).
- 9. Anton Vakhrushev, Alexander Ryzhkov, Maxim Savchenko, Dmitry Simakov, Rinchin Damdinov и Alexander Tuzhilin. "Lightautoml: Automl solution for a large financial services ecosystem". B: arXiv preprint arXiv:2109.01528 (2021).
- 10. Lucas Zimmer, Marius Lindauer и Frank Hutter. Auto-PyTorch Tabular: Multi-Fidelity MetaLearni for Efficient and Robust AutoDL. 2021. arXiv: 2006.13799 [cs.LG].
- 11. Davide Chicco и Giuseppe Jurman. Sepsis Survival Minimal Clinical Records. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C53C8N. 2023.

- 12. Ron Kohavi. *Census Income*. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5 1996.
- 13. The CDC. CDC Diabetes Health Indicators. DOI: https://doi.org/10.24432/C53919. URL: https://www.archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators (дата обр. 05.04.2023).
- 14. Wagner Dennis, Heider D. и Hattab Georges. Secondary Mushroom. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5FP5Q. 2023.
- 15. Mushroom. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5959T. 1987.
- Dirk Merkel. "Docker: lightweight linux containers for consistent development and deployment".B: Linux journal 2014.239 (2014), c. 2.
- 17. Anna Veronika Dorogush, Vasily Ershov и Andrey Gulin. "CatBoost: gradient boosting with categorical features support". B: arXiv preprint arXiv:1810.11363 (2018).
- 18. Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta и Masanori Koyama. "Optuna: A next-generation hyperparameter optimization framework". B: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019, c. 2623—2631.
- 19. Anaconda Software Distribution. Bep. Vers. 2-2.4.0. 2020. URL: https://docs.anaconda.com/.
- 20. Lukas Biewald. Experiment Tracking with Weights and Biases. Software available from wandb.com. 2020. URL: https://www.wandb.com/.
- 21. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg и др. "Scikit-learn: Machine learning in Python". B: the Journal of machine Learning research 12 (2011), с. 2825—2830.
- 22. Wes McKinney и др. "pandas: a foundational Python library for data analysis and statistics". B: Python for high performance and scientific computing 14.9 (2011), с. 1—9.
- 23. Stefan Van Der Walt, S Chris Colbert и Gael Varoquaux. "The NumPy array: a structure for efficient numerical computation". B: Computing in science & engineering 13.2 (2011), c. 22—30.
- 24. J. D. Hunter. "Matplotlib: A 2D graphics environment". B: Computing in Science & Engineering 9.3 (2007), c. 90—95. DOI: 10.1109/MCSE.2007.55.

25. Michael L. Waskom. "seaborn: statistical data visualization". B: Journal of Open Source Software 6.60 (2021), c. 3021. DOI: 10.21105/joss.03021. URL: https://doi.org/10.21105/joss.03021.

# Приложения

## А Диаграммы размаха для проведенных экспериментов

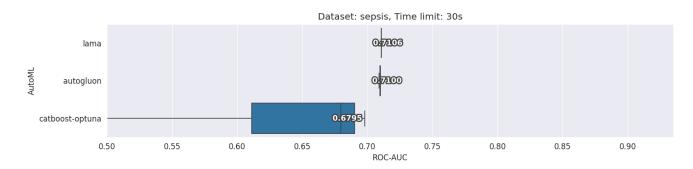


Рис. А.1: Ограничение 30 секунд, датасет Sepsis Survival Minimal Clinical Records

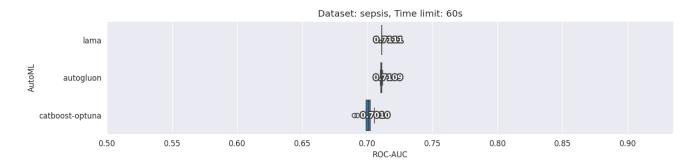


Рис. А.2: Ограничение 60 секунд, датасет Sepsis Survival Minimal Clinical Records

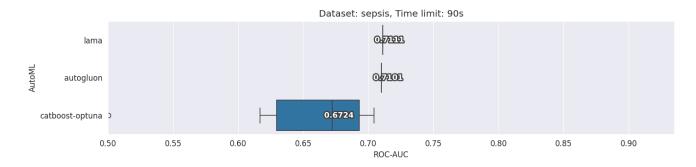


Рис. А.3: Ограничение 90 секунд, датасет Sepsis Survival Minimal Clinical Records

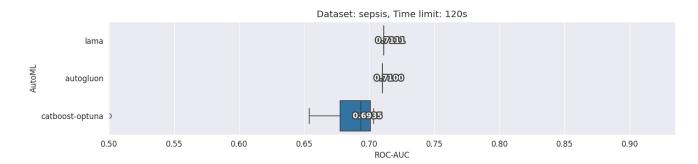


Рис. А.4: Ограничение 120 секунд, датасет Sepsis Survival Minimal Clinical Records

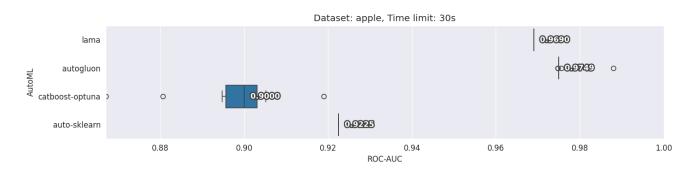


Рис. А.5: Ограничение 30 секунд, датасет Apple Quality

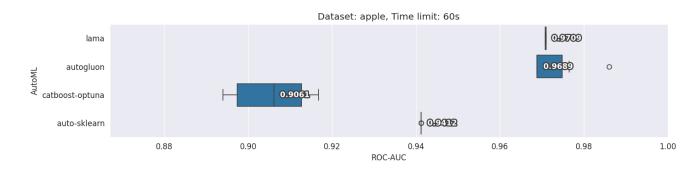


Рис. А.6: Ограничение 60 секунд, датасет Apple Quality



Рис. А.7: Ограничение 90 секунд, датасет Apple Quality

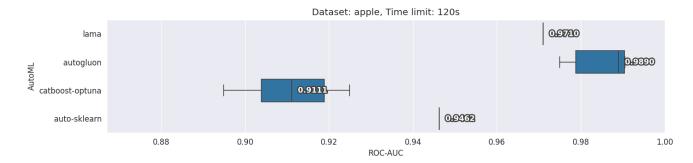


Рис. А.8: Ограничение 120 секунд, датасет Apple Quality

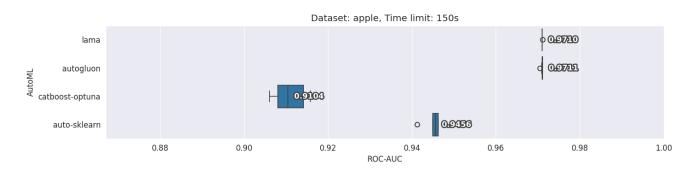


Рис. А.9: Ограничение 150 секунд, датасет Apple Quality

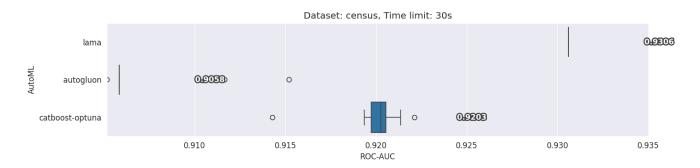


Рис. А.10: Ограничение 30 секунд, датасет Census Income

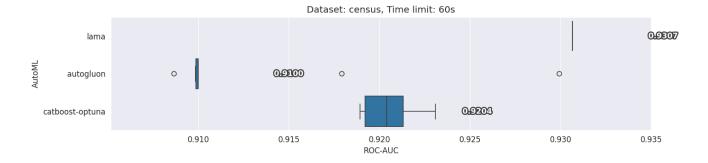


Рис. А.11: Ограничение 60 секунд, датасет Census Income

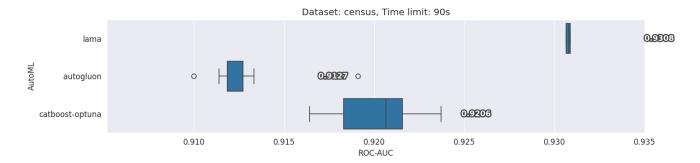


Рис. А.12: Ограничение 90 секунд, датасет Census Income



Рис. А.13: Ограничение 120 секунд, датасет Census Income

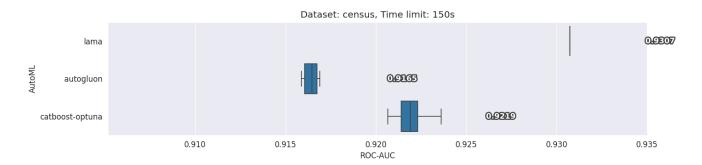


Рис. А.14: Ограничение 150 секунд, датасет Census Income

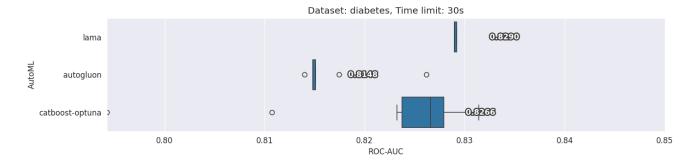


Рис. А.15: Ограничение 30 секунд, датасет CDC Diabetes Health Indicators

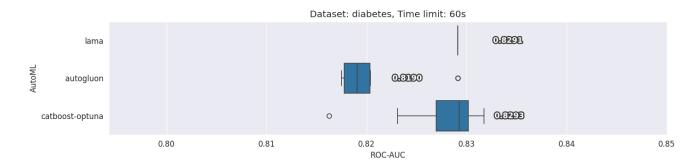


Рис. А.16: Ограничение 60 секунд, датасет CDC Diabetes Health Indicators

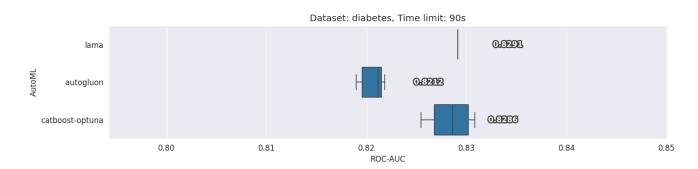


Рис. А.17: Ограничение 90 секунд, датасет CDC Diabetes Health Indicators

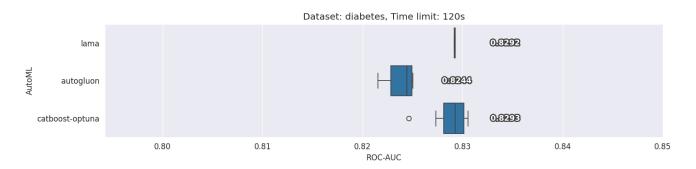


Рис. А.18: Ограничение 120 секунд, датасет CDC Diabetes Health Indicators

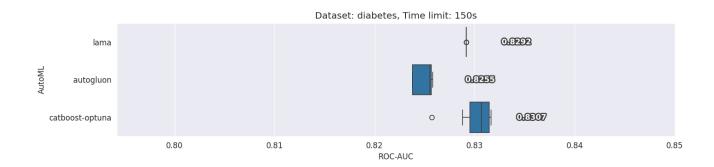


Рис. А.19: Ограничение 150 секунд, датасет CDC Diabetes Health Indicators

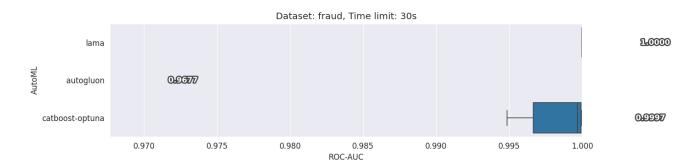


Рис. A.20: Ограничение 30 секунд, датасет Credit Card Fraud Detection Dataset 2023

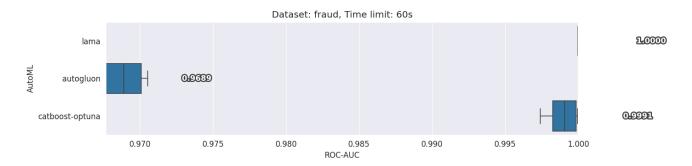


Рис. A.21: Ограничение 60 секунд, датасет Credit Card Fraud Detection Dataset 2023

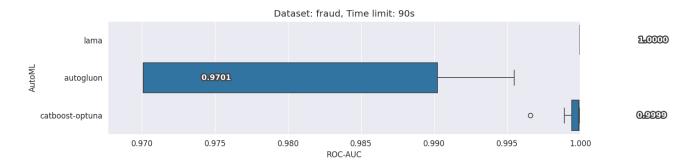


Рис. A.22: Ограничение 90 секунд, датасет Credit Card Fraud Detection Dataset 2023