



# Технологии AutoML и MLOps в Loginom для выпускных студенческих работ

Николай Паклин, тимлид Loginom Company

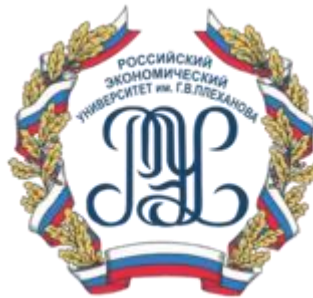
В академической  
программе Loginom на  
01.10.2024

# 110 вузов

140+ преподавателей

4700+ студентов и 100+ ВКР

в 2023/2024 учебном году



## Признаки хорошей ВКР по аналитике данных

- Использование State-of-Art методов и подходов
- Практическая значимость и законченность работы
- Современный стек инструментов и технологий
- Научность (для магистров)

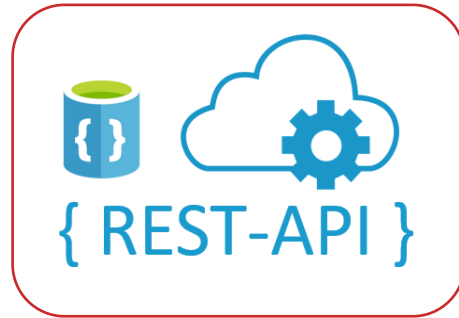
## Признаки не очень хорошей ВКР по аналитике данных

- Использование только встроенных в Loginom алгоритмов анализа, когда задача требует других алгоритмов
- Построенная модель не готова к внедрению
- Плохо проработаны вопросы интеграции Loginom с первичными источниками данных
- Слабое использование внешних инструментов и сервисов (Github, REST, порталы открытых данных)
- Отсутствует воспроизводимость исследования

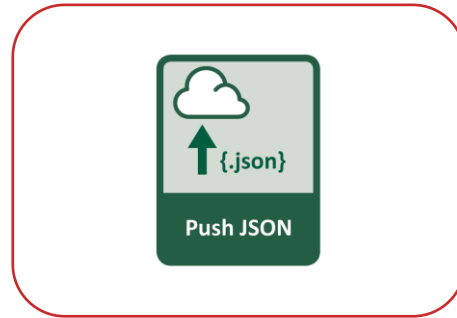


Loginom – low-code платформа для аналитики самообслуживания, но законченная работа требует использования стека из нескольких систем и технологий.

# Доступные способы повышения уровня выпускных работ



Обогащение  
данных по REST API



Публикация веб-  
сервиса



Репозиторий Git

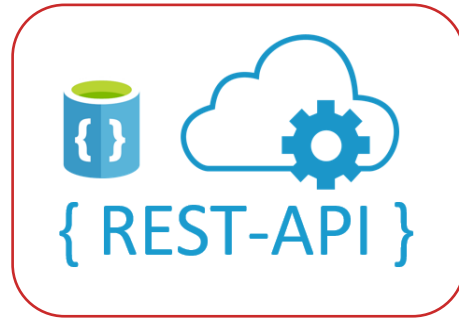


Модели из Python

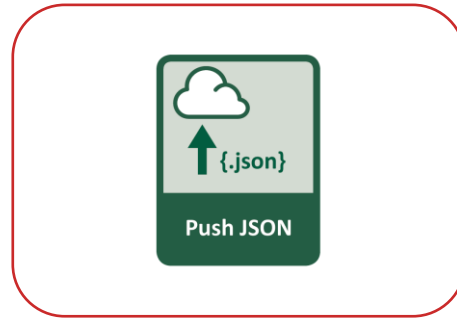


Фреймворки для  
MLOps

# Доступные способы повышения уровня выпускных работ



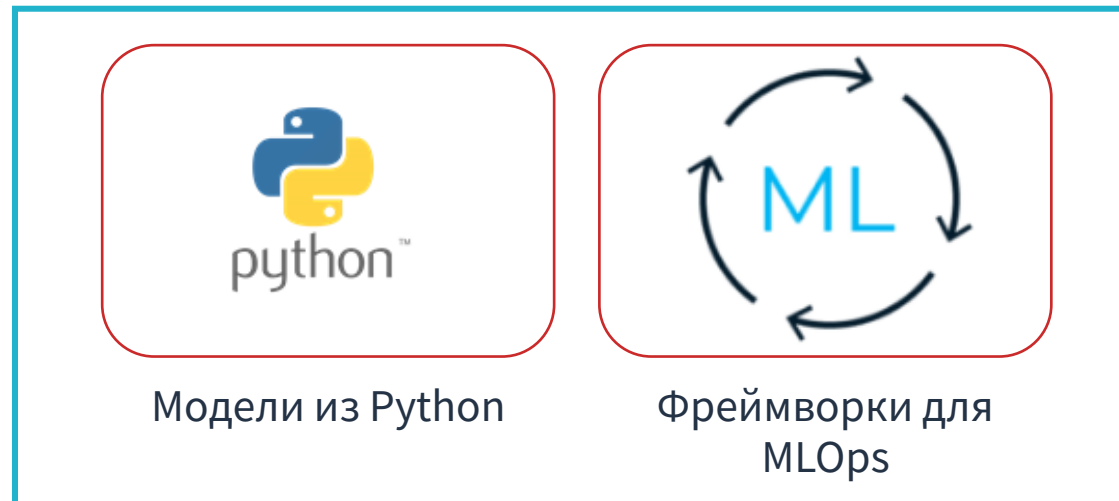
Обогащение данных по REST API



Публикация веб-сервиса



Репозиторий Git



Модели из Python

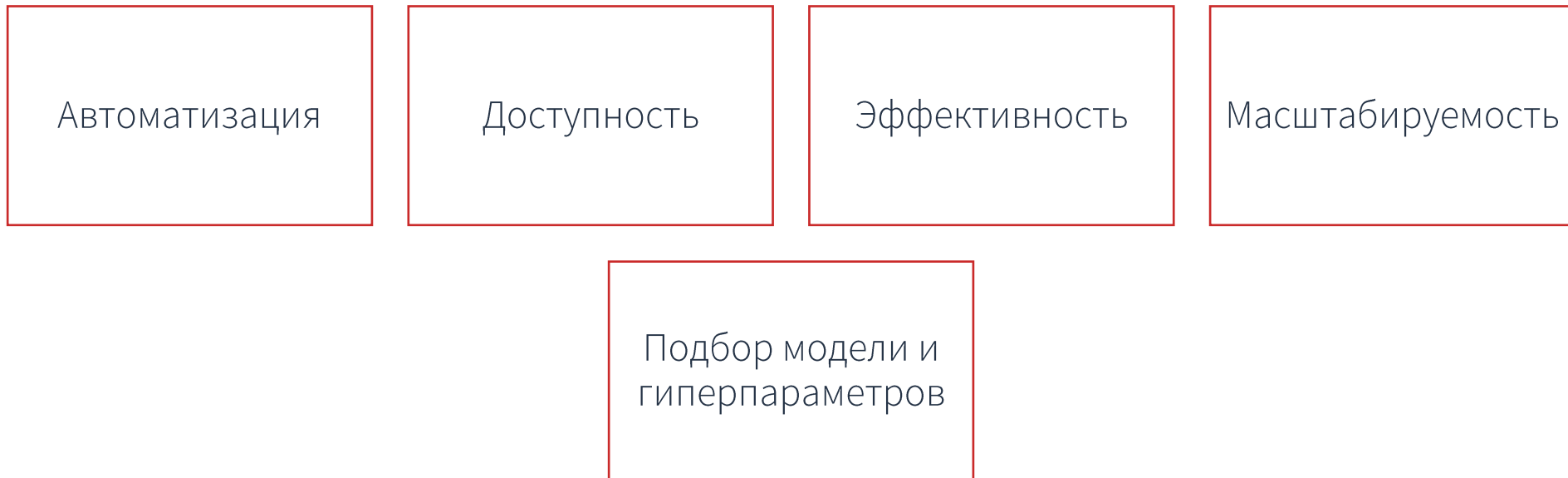
Фреймворки для MLOps

# Технология AutoML в Loginom

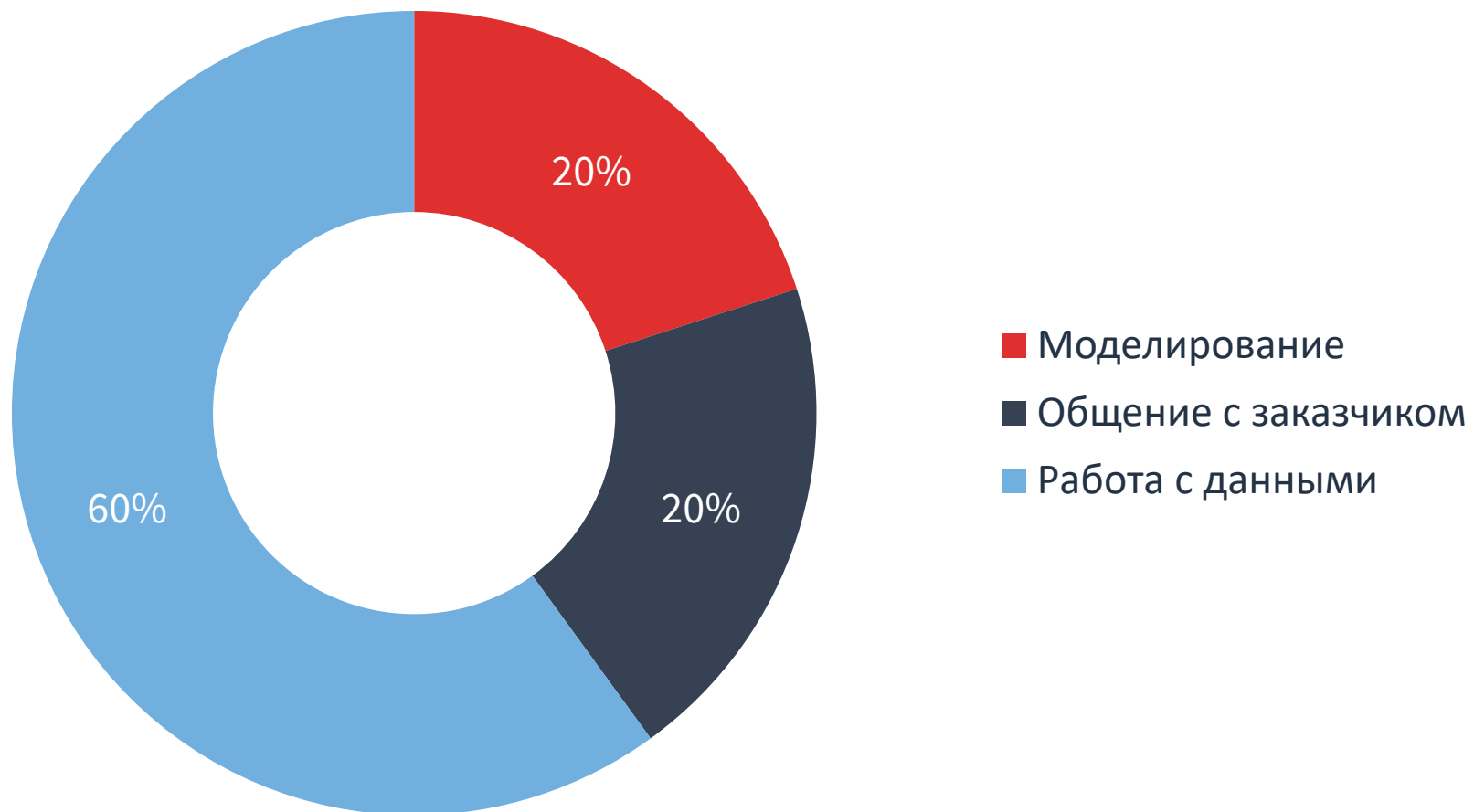


## Концепция AutoML

Процесс автоматизации создания и применения моделей машинного обучения без необходимости глубоких знаний в этой области.



## На что тратит время аналитик данных?



# Общий принцип работы AutoML

Цель – автоматизировать обучение моделей на табличных данных.

## Загрузка данных

Считаем, что признаки и таргеты уже собраны пользователем

1



## Обработка данных

Очистка от плохих фичей, автокодирование

2



## Обучение и тюнинг моделей

Набор алгоритмов произвольный, в основном -бустинги

3

Catboost

LightGBM

...

## Применение

Итоговая модель – ансамбль, blending. Предсказание на ретро-выборке

4



# Дополнительные компоненты



?

**Интерпретация**



?

**Отбор признаков**



?

**Отчетность**

- **Дополнительный функционал варьируется**
- **Часто – лишь номинальная реализация**

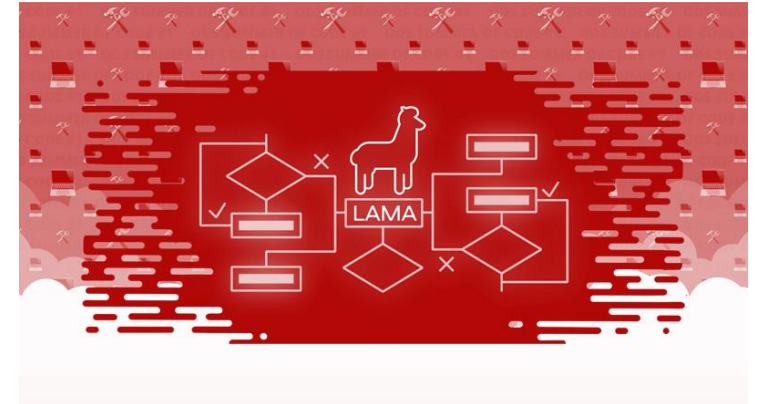
## Популярные python-библиотеки

- LightAutoML (LAMA) от Сбера
- AutoGluon
- H2O AutoML
- Auto-SkLearn
- FEDOT

## LightAutoML (LAMA)

- Отечественная разработка
- Высокое качество моделей
- Гибкая конфигурация обучения
- Полноценная интерпретация (SHAP)

**!** Один из самых практичных AutoML  
(мнение экспертов отрасли)



## Проблемы AutoML

- Малое покрытие автоматизации
- Низкое доверие пользователей
- Технические проблемы
- Работа с большими данными

# Управление моделями и экспериментами

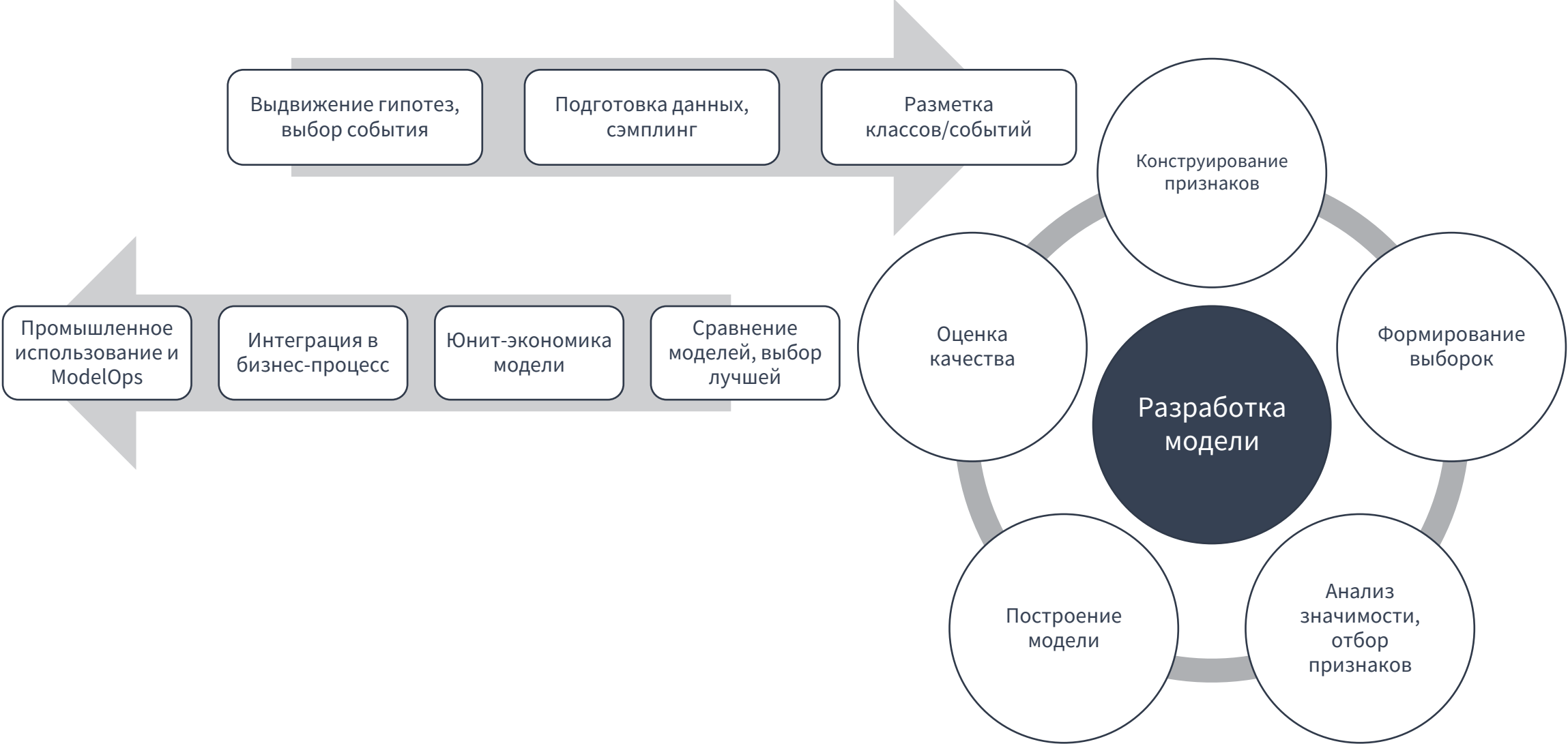


## Операционализация аналитики данных

- **Воспроизводимость эксперимента** – мера вероятности того, что, получив один результат эксперимента, вы сможете провести тот же эксперимент с теми же параметрами и получить точно такой же результат.
- **ModelOps** – набор практик, направленных на унификацию процессов разработки и развертывания аналитических моделей и решений

*(часто идет речь о моделях machine learning - MLOps)*

# Общая схема процесса построения ML-модели



## Требования к фреймворку

- Бесплатный в облаке, возможно, с ограничениями
- Работа в локальной сети
- Поддержка сторонних файловых хранилищ
- Подходит для систем класса low-code
- Низкий порог входа для аналитика
- Богатый Rest API



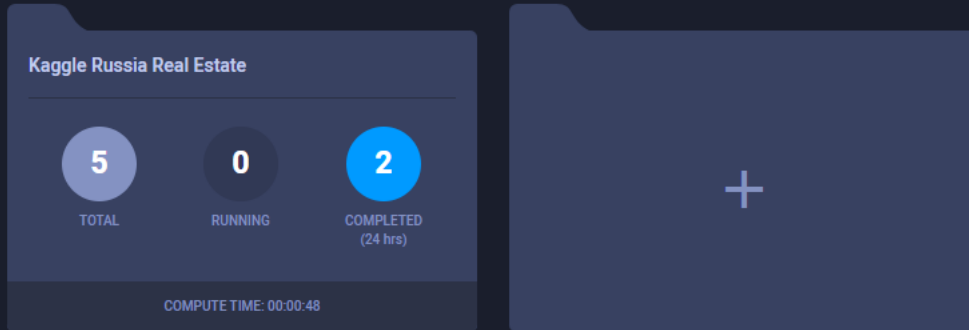
[www.clear.ml](http://www.clear.ml)

# ClearML - Модули для трекинга экспериментов





### RECENT PROJECTS [VIEW ALL](#)



### RECENT EXPERIMENTS

[MANAGE WORKERS AND QUEUES](#)

TYPE	TITLE	PROJECT	STARTED	UPDATED	STATUS
Training	XGBoost Regressor (15 признаков)	Kaggle Russia Real Estate	May 3 2024 16:41	May 3 2024 18:12	Published
Training	XGBoost Regressor (15 признаков)	Kaggle Russia Real Estate	May 3 2024 16:28	May 3 2024 16:28	Failed
Training	Линейная регрессия (15 признаков)	Kaggle Russia Real Estate	May 3 2024 15:05	May 3 2024 15:05	Completed
Data Processing	Kaggle-Russia-Real-Estate	Kaggle Russia Real Estate/.datasets/Kaggle-Russia-Real-Estate	May 3 2024 14:22	May 3 2024 14:23	Completed
Training	Линейная регрессия (14 признаков)	Kaggle Russia Real Estate	Apr 27 2024 17:10	Apr 27 2024 17:11	Completed

# Веб-интерфейс ClearML: проекты и последние эксперименты

[Get Started](#) <sup>3</sup>





+ NEW EXPERIMENT

OPEN ARCHIVE



EXPERIMENTS LIST

SORTED BY

Linear regression (15 features) Completed

с геофичей Санкт-Петербург и область

Updated 10 days ago Created by Nikolay Paklin

Linear regression (14 features) Completed

Санкт-Петербург и область

Updated 16 days ago Created by Nikolay Paklin

Linear regression (14 features) Failed

Санкт-Петербург и область

Updated 16 days ago Created by Nikolay Paklin

COMPLETED

Linear regression (15 features) ID d2116cda...

с геофичей Санкт-Петербург и область

EXECUTION CONFIGURATION ARTIFACTS INFO CONSOLE SCALARS PLOTS

USER PROPERTIES

Properties	USER PROPERTIES
	Версия Loginom 7.1.5
	Редакция Loginom Standard

HYPERPARAMETERS

Датасет

Датасет	ДАТАСЕТ
Датасет Id	0b107a9657854cb9bb950355f0950715
Имя датасета	Kaggle-Russia-Real-Estate

Модель

Модель	МОДЕЛЬ
Алгоритм	Линейная регрессия
Нормализация	False
Регуляризация	Ridge

Сводка

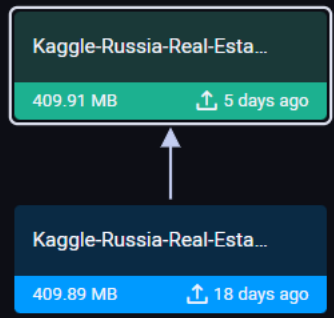
Карточка эксперимента в ClearML: версия ПО, артефакты (модели, сценарии, код) и гиперпараметры

OPEN ARCHIVE

VERSIONS LIST SORTED BY

- Kaggle-Russia-Real-Estate v 1.0.2** Final
  - preprocessed с геофичей
  - Updated 5 days ago Created by Nikolay Paklin
- Kaggle-Russia-Real-Estate v 1.0.1** Final
  - preprocessed
  - Updated 18 days ago Created by Nikolay Paklin
- Kaggle-Russia-Real-Estate v 1.0.0** Final
  - original
  - Updated 18 days ago Created by Nikolay Paklin

DETAILS



**VERSION INFO**

Kaggle-Russia-Real-Estate v1.0.2  
**FINAL**

ID 0b107a96...

Parent -

Size 409.91 MB (original) / 130.85 MB (compressed)

File count 4

Link count 0

**FILES CHANGED**

Added 0

Modified 1

Removed 0

Size 20.35 KB

Kaggle-Russia-Real-Estate v1.0.2

CONTENT PREVIEW CONSOLE

File Name (4 files)	File Size (total 409.91 MB)	Hash (SHA2)
raw-data/all_v2.csv	389.27 MB	46390bd7f9ac41375e7151f9691e83910c9621f...
raw-data/all_v2_1region.lgd	20.58 MB	ac7a2370361872b2c0771280b19b4d945ad6a9...
raw-data/data-prep.lgp	64.84 KB	63cf48e0ac753ec73753be7bceec7b5f7377a9c1...
raw-data/russia-real-estate-20182021_link.txt	71 B	ec552cafe63fcee88a535a25bde2673b81d0110...

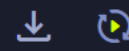
Датасеты наследуются друг от друга, имеют версии и тэги





+ NEW EXPERIMENT

OPEN ARCHIVE



EXPERIMENTS LIST SORTED BY

Linear regression (15 признаков) Completed  
с геофичей Санкт-Петербург и область  
Updated 10 days ago Created by Nikolay Paklin

Linear regression (14 признаков) Completed  
Санкт-Петербург и область  
Updated 16 days ago Created by Nikolay Paklin

Linear regression (14 признаков) Failed  
Санкт-Петербург и область  
Updated 16 days ago Created by Nikolay Paklin

COMPLETED

Linear regression (15 признаков) ID d2116cda...

с геофичей Санкт-Петербург и область

EXECUTION CONFIGURATION ARTIFACTS INFO CONSOLE SCALARS PLOTS DEBUG SAMPLES

Summary			
R <sup>2</sup> (train)	MAPE (train)	R <sup>2</sup> (test)	MAPE (test)
0.7191002859318962	0.2139850207170322	0.643921859032758	0.20300311764034515

Карточка эксперимента в ClearML: метрики







+ NEW EXPERIMENT

OPEN ARCHIVE



EXPERIMENTS LIST

SORTED BY

Linear regression (15 features) ✓ Completed

с геофичей Санкт-Петербург и область

Updated 10 days ago • Created by Nikolay Paklin

Linear regression (14 features) ✓ Completed

Санкт-Петербург и область

Updated 16 days ago • Created by Nikolay Paklin

Linear regression (14 features) ✗ Failed

Санкт-Петербург и область

Updated 16 days ago • Created by Nikolay Paklin

Карточка эксперимента в ClearML: диаграмма рассеяния

COMPLETED



Линейная регрессия (15 признаков)

ID d2116cda...

с геофичей Санкт-Петербург и область

EXECUTION

CONFIGURATION

ARTIFACTS

INFO

CONSOLE

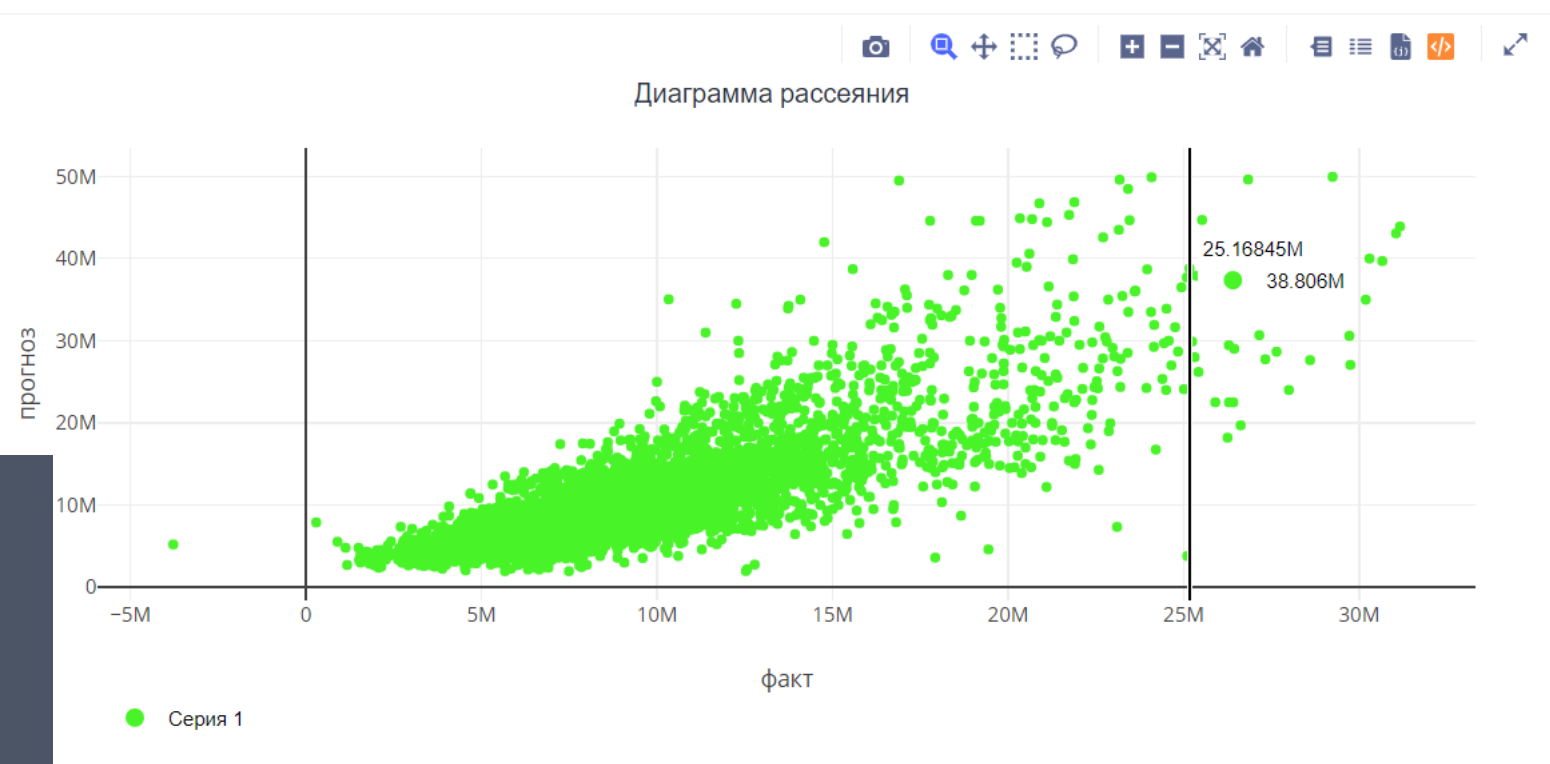
SCALARS

PLOTS

DEBUG SAMPLES



Диаграмма рассеяния



## Плюсы использования в ВКР

### AutoML:

- экономия времени
- высокое качество ML-моделей
- использование последних достижений в области ML
- достоверность полученных результатов
- низкий порог входа (при наличии интеграции с Loginom)

### MLOps:

- хранение истории экспериментов
- воспроизводимость экспериментов
- низкий порог входа (при наличии интеграции с Loginom)

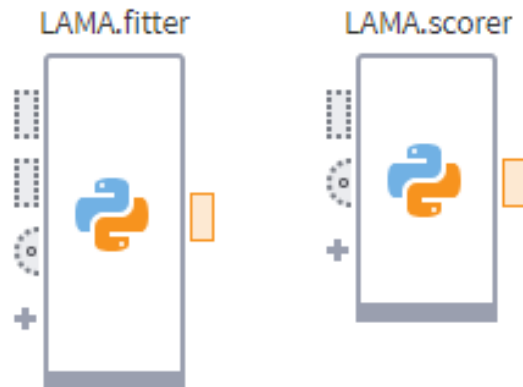
# Реализация в Logiном

## Основные принципы

- Соответствие парадигме low-code
- Быстрый старт
- Шаблоны сценариев с предустановленными действиями

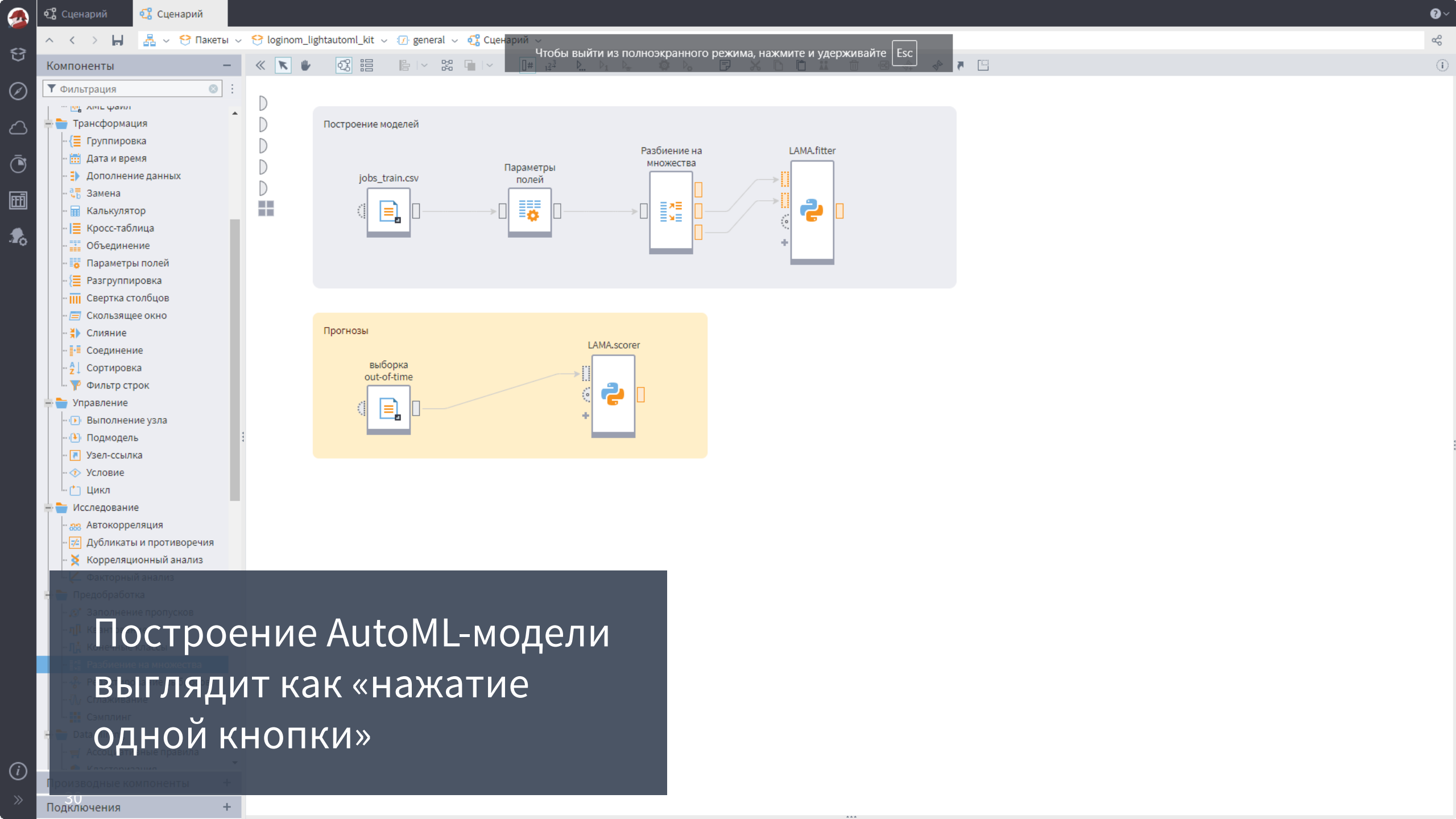
# Реализация интеграции Loginom с LightAutoML

- Библиотека компонентов **LightAutoML Kit**
- Все настройки задаются в переменных узлов
- Итоговая модель и отчет формируются в отдельной папке



Настройка переменных

Метка	Имя	Назначение	Значение
ab	ОБЩИЕ ПЕРЕМЕННЫЕ	XXX	Не задано
ab	Виртуальное окружение	virt_env	Не задано
0/1	Всегда создавать новую п...	new_folder	Не задано
ab	Библиотека для сериализ...	serialize_lib	Не задано
ab	Имя файла	model_filename	Не задано
ab	Поле с идентификатором	id_fieldname	Не задано
ab	Выходное поле с разбиен...	partition_fieldname	Не задано
ab	Целевое поле	target_field	Не задано
ab	Игнорируемые поля	ignore_fields	Не задано
0/1	Полный выходной набор	full_output_dataset	Не задано
0/1	Сохранять метки полей	keep_labels	Не задано
ab	Словарь других ролей	role_dict	Не задано
ab	КЛАСС TabularAutoML	XXXX	Не задано
ab	task	task	Не задано
12	timeout	timeout	Не задано
12	cpu_limit	cpu_limit	Не задано
ab	reader_params	reader_params	Не задано

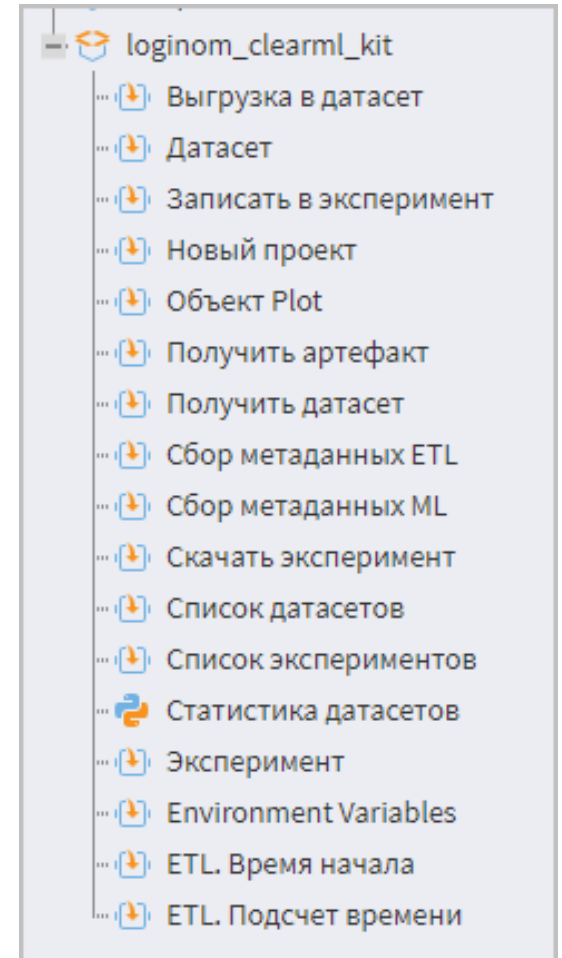


Построение AutoML-модели  
выглядит как «нажатие  
одной кнопки»

## Состав интеграции Loginom с ClearML

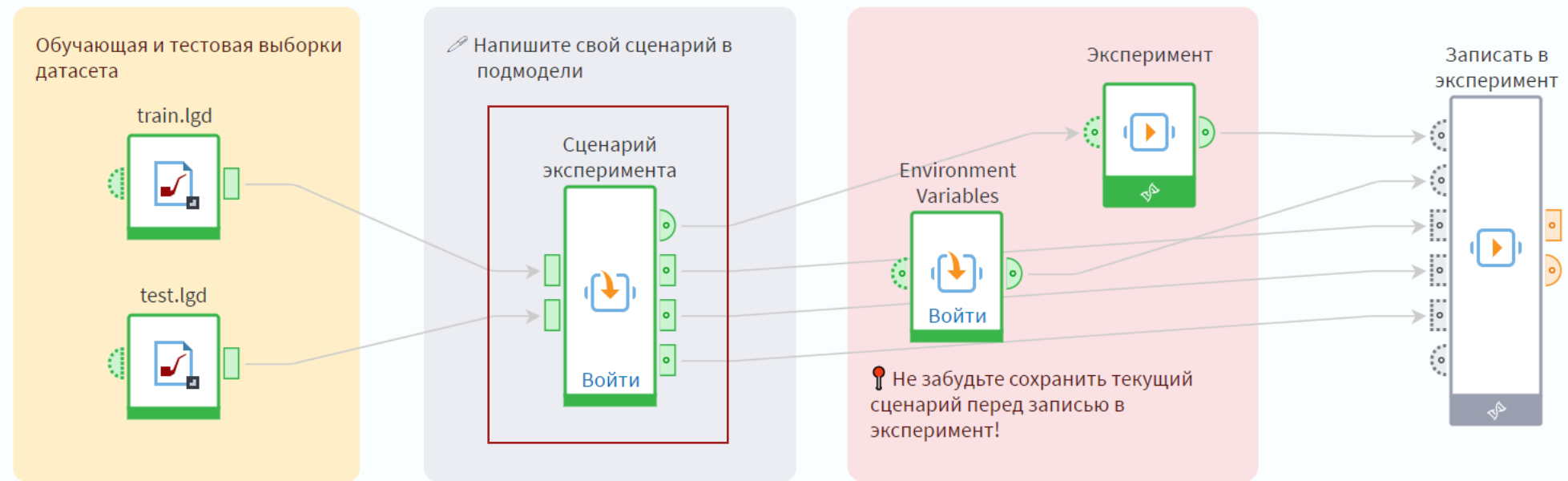
1. Библиотека компонентов **ClearML Kit**
2. Шаблоны сценариев

Для интеграции используется python-библиотека **clearml**



Напишите сценарий эксперимента, используя файлы выборок train и test из папки samples.

Сохраните сценарий и метаданные в эксперимент.



Аналитик наполняет узлами подмодель **Сценарий эксперимента**. Ряд показателей (версия Loginot, настройки узла-модели и другие) собирается автоматически



# СКОРО

- Релиз библиотеки компонентов **LightAutoML Kit**
- Релиз библиотеки компонентов **ClearML Kit**
- Серия вебинаров в Мастерской Loginom Skills



Проект «Мастерская»

<https://mastering.loginom.ru>

Библиотеки компонентов –

<https://marketplace.loginom.ru/libraries>

Telegram-канал Skills –

[https://t.me/loginom\\_skills](https://t.me/loginom_skills)