



Low-code на автопилоте: AutoML и автодокументирование в Loginom

Николай Паклин



- Николай Паклин – тимлид группы «Обучение»
- 20 лет в low-code аналитике и Loginom Company
- С командой запустил Loginom Skills, проект «Мастерская», библиотеки компонентов для Loginom
- Бизнес-тренер, преподаватель, автор книг и учебников по аналитике данных

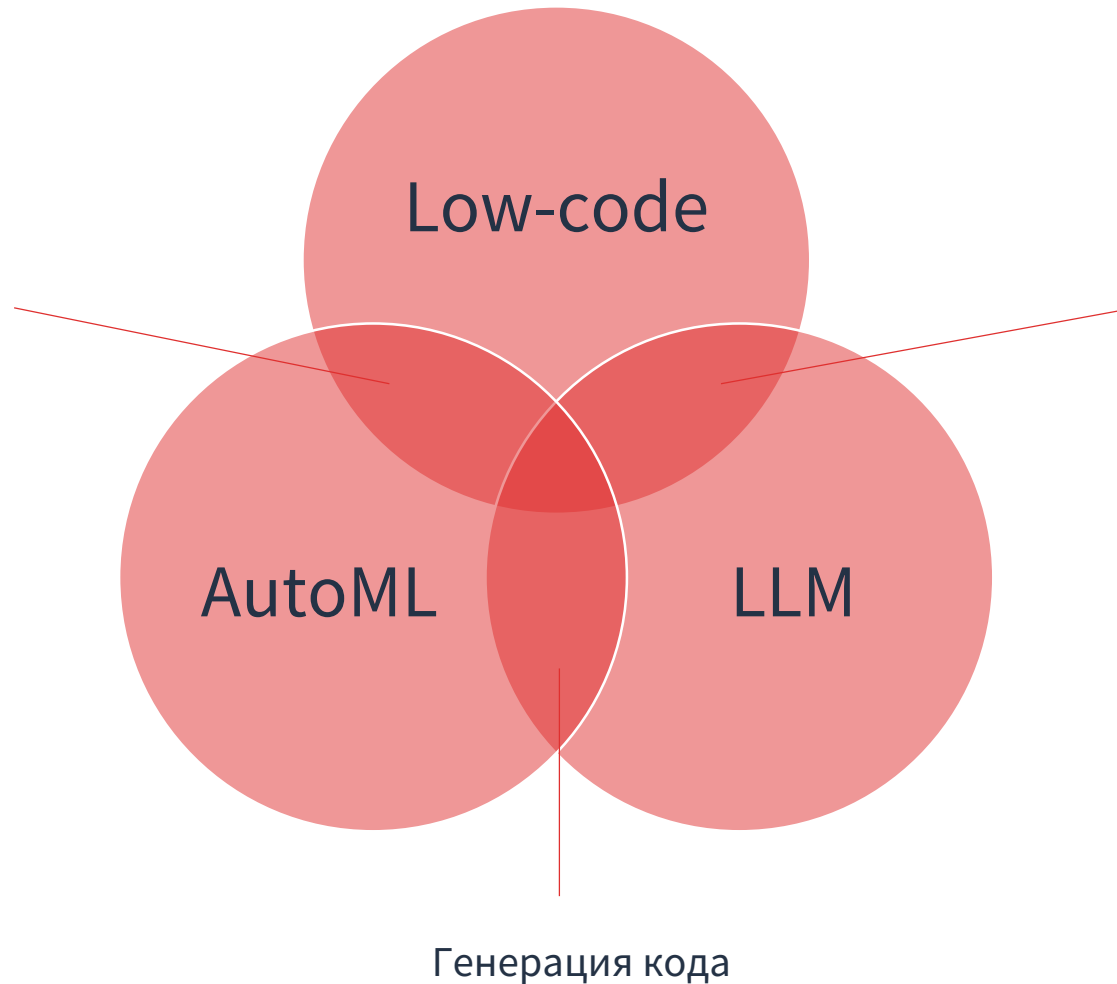
О чем будем говорить

1. Технологии аналитики самообслуживания
2. Автоматизация машинного обучения
3. Интеграция технологий искусственного интеллекта в обработку данных
4. Автодокументирование сценариев с помощью нейросетей

AutoML в Loginom

Синергия технологий аналитики самообслуживания

- Усиление функциональности
- Эффективные ML-модели без кодинга



- Генерация workflow, скриптов
- Очистка и обогащение данных
- Интерпретация результатов
- Творческие задачи
- ...

AutoML – демократизация машинного обучения

- 1. AutoML – это автоматизация полного цикла машинного обучения: от подготовки данных до развертывания моделей.**
- 2. Не замена дата-аналитика, а мощный инструмент для повышения эффективности и расширения возможностей.**

Классический подход (ML)

- Глубокое знание алгоритмов ML
- Ручная настройка гиперпараметров
- Сложный процесс разработки и развертывания
- Высокий порог вхождения
- Требуется экспертов Data Science
- Основной фокус: алгоритмы и кодирование

Современный подход (AutoML)

- Автоматизированный выбор и настройка алгоритмов
- Ускоренная разработка моделей
- Более простой процесс развертывания
- Доступность для широкого круга аналитиков
- Фокус на бизнес-задаче и 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



Фреймворки AutoML

- 1. Текущие открытые фреймворки AutoML отлично решают задачу моделирования.**
- 2. Есть ряд направлений и слабых мест для улучшения.**

Проблемы:

- малое покрытие автоматизации;
- низкое доверие пользователей;
- технические проблемы;
- большие данные.



Фреймворк LightAutoML от Сбера

- 1. Один из самых практичных AutoML (мнение экспертов Alfa Digital).**
- 2. Интеграция с Loginom в рамках библиотеки компонентов Python Kits.**

Плюсы:

- Высокое качество моделей
- Гибкая конфигурация обучения
- Автоматический отчет
- Полноценная интерпретация (SHAP)
- Быстрый и долгий режимы

Минусы

- Медленный отбор признаков
- Неполная документация (имеется русскоязычный тг-чат)

Первый релиз – май 2021.

Сценарий

Сценарий

Пакеты

otp_bank

Модуль1

Сценарий

Компоненты

Производные компоненты

Фильтрация

Текущий модуль

otp_bank

loginom_lightautoml_kit

LAMA.fitter

LAMA.scorer

Train

ОТП банк

15 223

Параметры полей

15 223

Параметры полей

14 910

Настройки

LAMA

17

LAMA.fitter

30 133

LAMA report

This report was generated automatically.

▼ Model overview

► Model parameters

▼ Summary results

Results for data samples:

Evaluation parameter	Validation sample	Test sample
AUC-score	0.7078	0.7040
Precision	0.2323	0.2273
Recall	0.4879	0.4583
F1-score	0.3147	0.3039

▼ Data overview

▼ Train data summary

Parameter	Value
Number of records	15223
Total number of features	52
Used features	49
Dropped features	3

► Train data details

► Feature importance

▼ Detailed model results

► Results on validation sample

► Results on test sample

Весь **AutoML** сводится к двум компонентам, для обучения модели и «прогона» через нее. В папке сценария формируется отчет и pkl-файл.

Подключения

AutoML: рутинизация задач машинного обучения за десятилетие

1. **Пример: задача предсказания отклика клиентов ОТП Банка**
2. **Конкурс проводился в 2011 году в рамках конференции ММРО-15.**

Традиционный подход (ML)

- Много узлов
- Предобработка данных
- Заполнение пропусков
- Построение моделей
- Несколько часов работы аналитика
- Метрика победителя ROC-AUC: **0,6935**

LightAutoML в Loginom

- 2 узла, не считая импорта
- Минимум настроек
- Результат – до 600 секунд
- Метрика ROC-AUC: **0,7040**

LLM в Loginom

LLM в аналитике данных: больше автоматизации, меньше рутины в анализе.

1. Обогащение данных
2. Очистка данных
3. Генерация новых признаков
4. ...

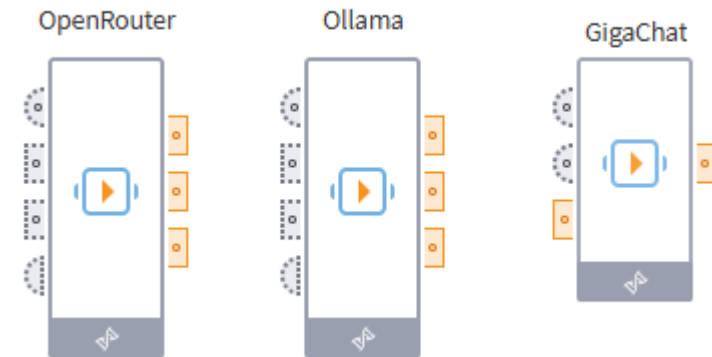
Компоненты интеграции

Интеграция Loginom с LLM:

1. OpenRouter
2. GigaChat
3. Ollama

Компоненты будут выпущены в рамках библиотеки **Loginom REST API Kit**.

- Задание промпта в переменных
- Парсинг json-ответа нейросети
- Кэширование результата



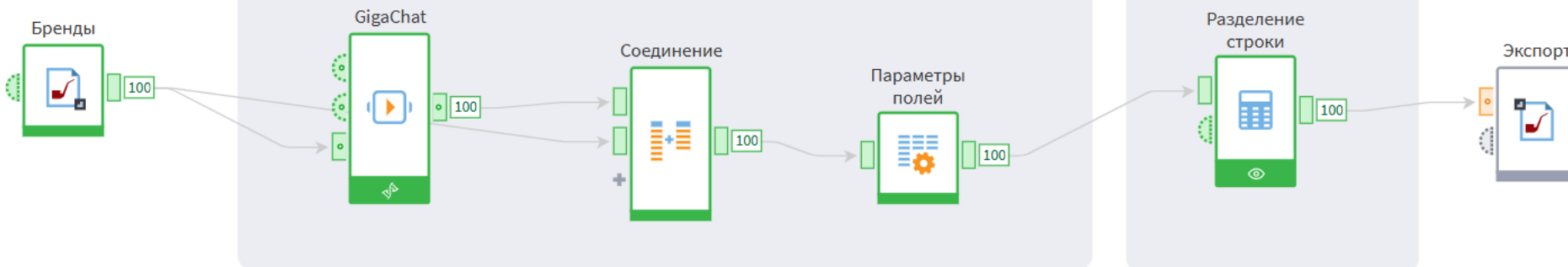


Пусть у нас есть список брендов в сети магазинов одежды и косметики класса лакшери.
Мы хотим дополнить наши данные страной бренда и кратким описанием продукции бренда.
Воспользуемся для этого генеративным ИИ - GigaChat через REST API.

ПРОМПТ:

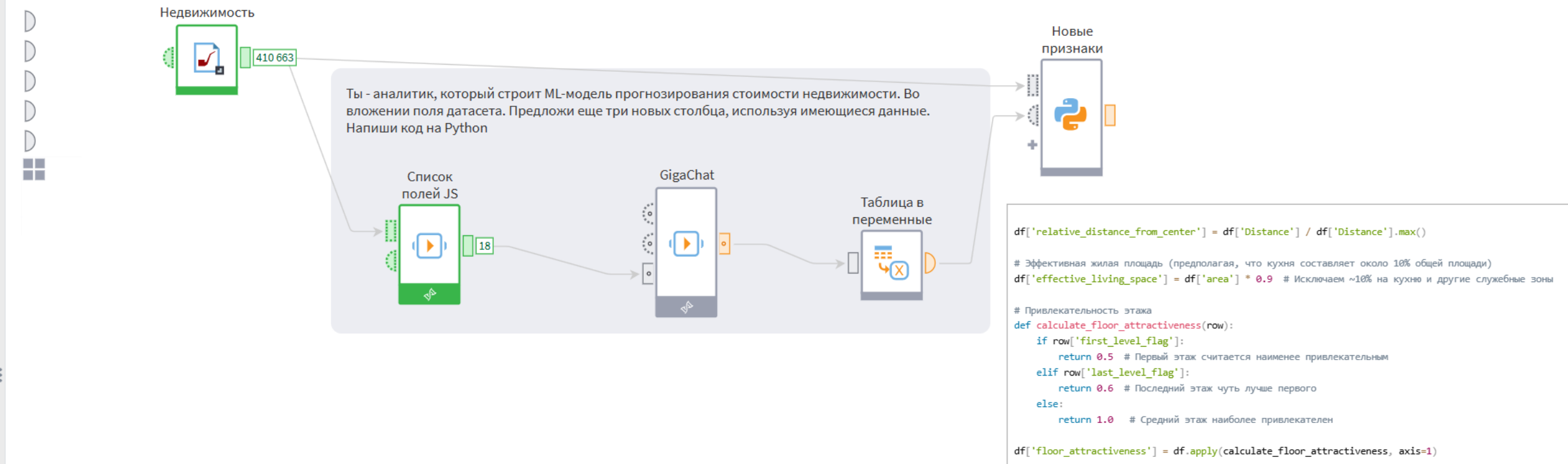
- 1) Ты аналитик торговой сети класса luxury. Тебе нужно обогатить данные по брендам из продаж
- 2) Ответ должен быть в формате "Название страны"|"Информация о бренде". В конце не ставь знаки препинания. Пустая страна недопустима
- 3) К какой стране относится данный бренд? Кратко о товарах бренда

В переменную GIGA_CHAT_AuthorizationKey нужно ввести код авторизации.
Получить бесплатно: <https://developers.sber.ru/portal/products/gigachat-api>



Пример: обогащение справочника брендов сети
«лакшери» страной и кратким описанием через
GigaChat

Результат



Пример: генерация новых признаков для задачи прогнозирования стоимости недвижимости с помощью **GigaChat** и **Python**

LLM для Loginot. Сервис автодокументирования

Автодокументирование сценариев

Автодокументирование – это **вызов** для low-code платформ, требующий специальных решений.

- Отсутствие стандартных текстовых описаний

Визуальные связи трудно автоматически преобразовать в понятный текст

- Многообразие компонентов

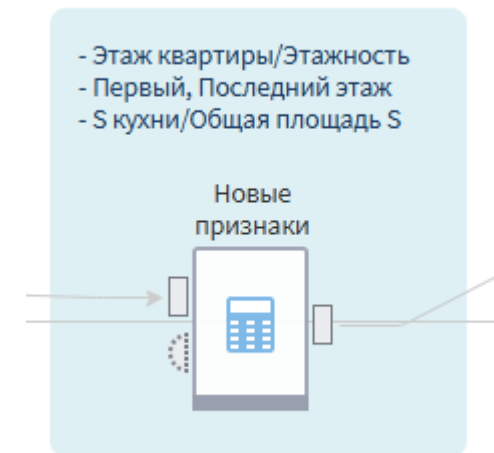
Разные блоки требуют индивидуальной обработки и описания

- Отсутствие метаданных

Сложности с извлечением параметров и связей без дополнительных усилий

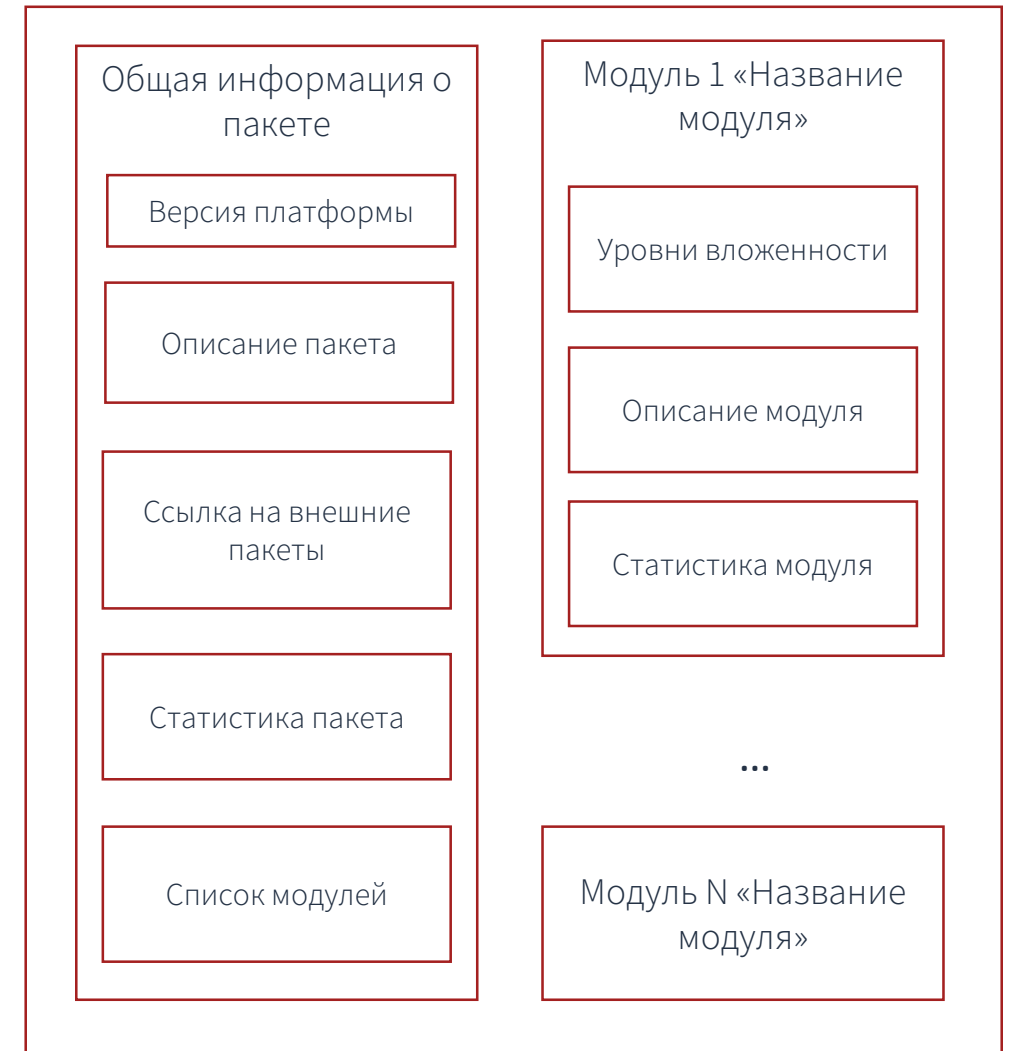
Пути решения

1. Использовать активнее заметки и комментарии: добавлять пояснения к элементам и связям
2. Задавать осмысленные метки узлам: правильно называть компоненты и связывать их смыслом
3. Использовать LLM: автоматический помощник для генерации описаний на основе комментариев и меток



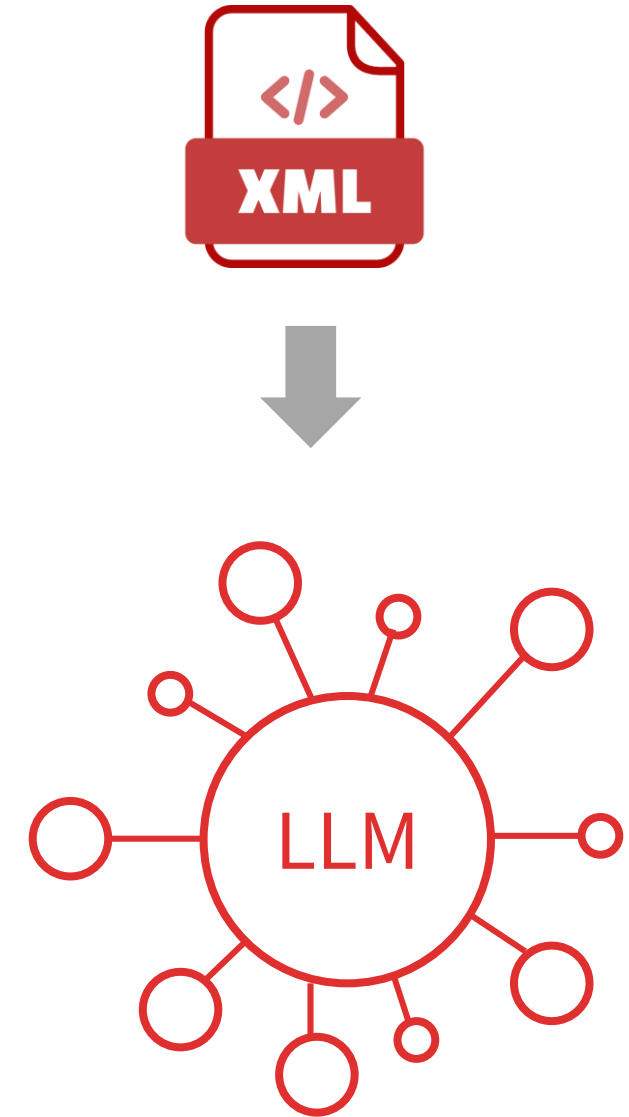
Каким мы видим идеальный отчет

1. Развернутое текстовое описание сценария каждого модуля, логики работы подмоделей
2. Статистика по каждому модулю
3. Резюме по пакету в целом
4. Уровни вложенности сценария
5. Ссылки на внешние пакеты (при наличии)



Проблемы «лобового» подхода

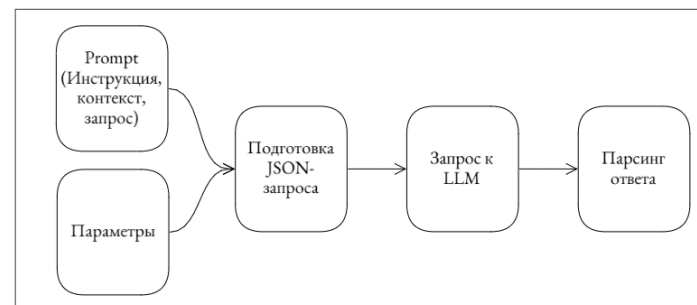
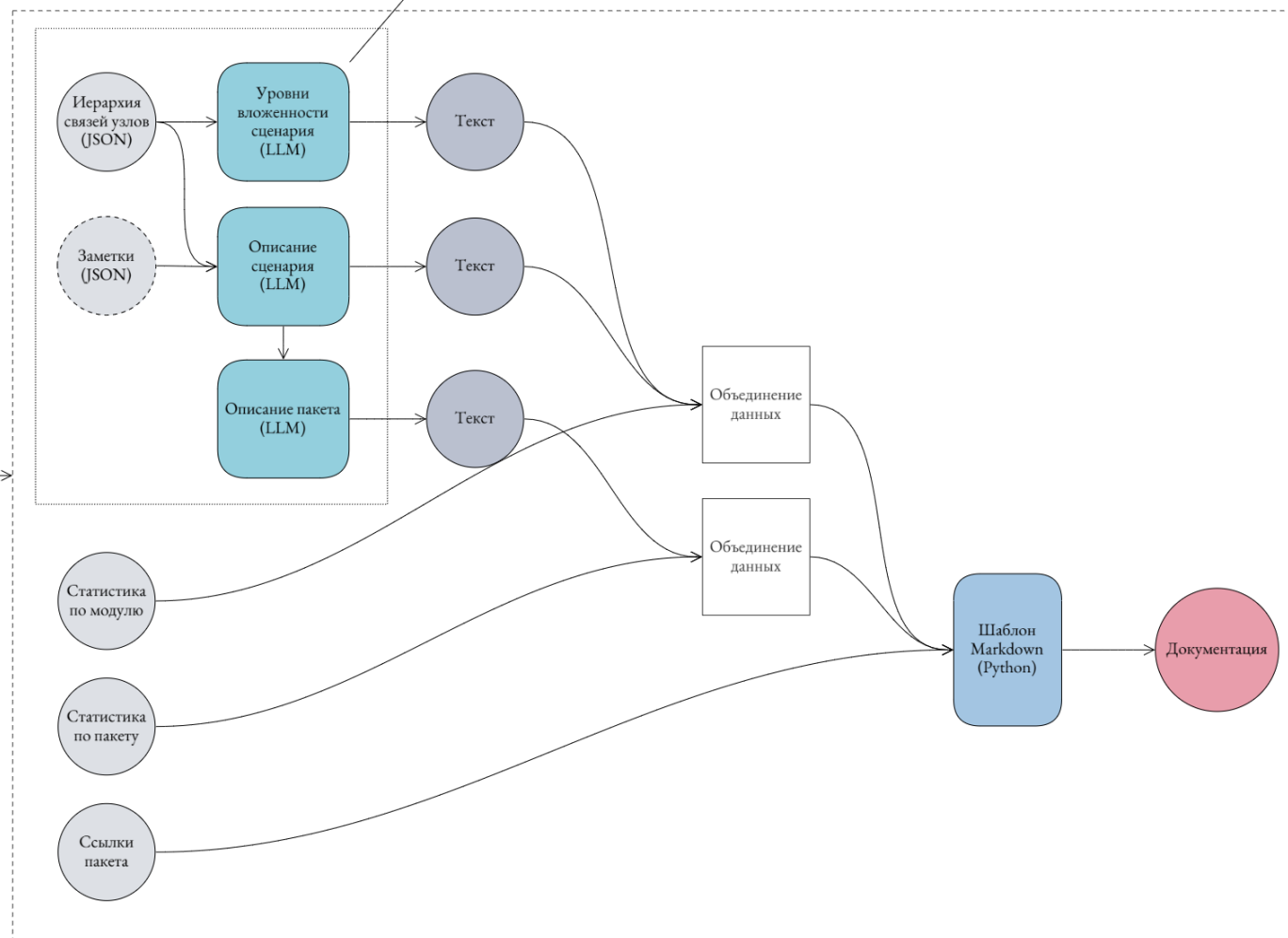
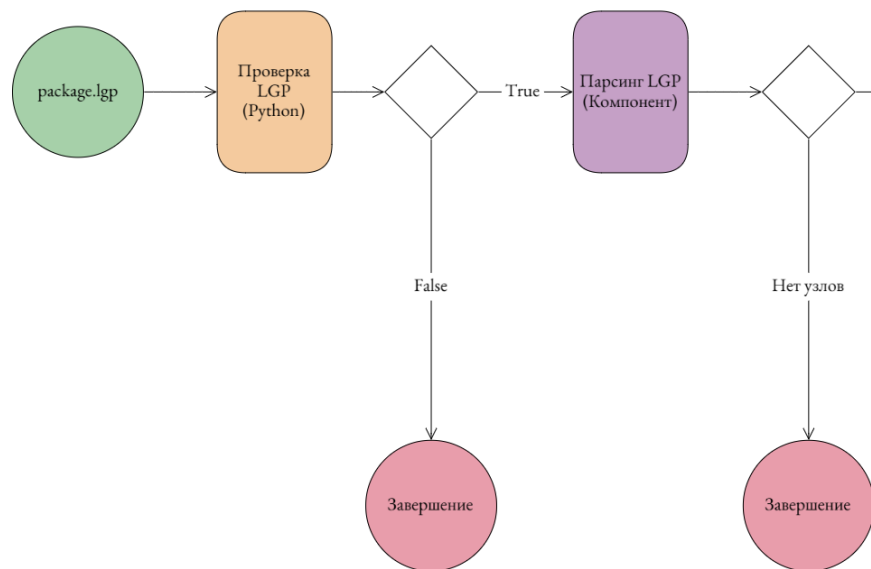
1. Большой объем служебных символов и избыточной информации в XML
2. Необходимость для модели самостоятельного определения иерархии и связей
3. Сложность правильной интерпретации инструкций
4. Низкое качество генерируемых отчетов



Эффективный подход

1. Набор правил
2. Предобработка информации
3. Четкие инструкции и параметры
4. Использование внешней LLM
(например, DeepSeek v. 3)
5. Использование шаблона отчета
6. Реализация - веб-сервис на Loginom!

Многоступенчатая схема дает нужный результат и экономит токены



AIDOC 1.0Интеллектуальная документация
для сценариев Loginom

Тестовый пример

[Скачать demo.lgp](#)[Посмотреть онлайн](#)

Сгенерируйте описание вашего сценария

① подключено

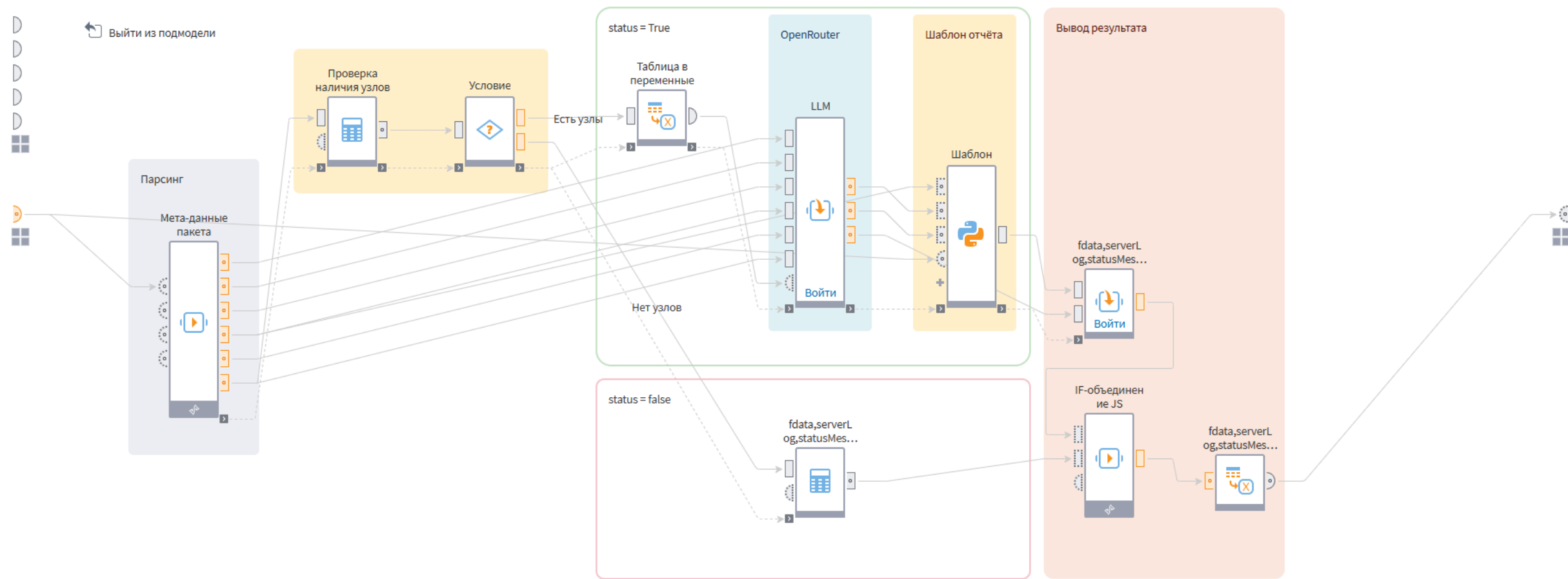
Перенесите файл сюда или [нажмите для загрузки](#)

Мы принимаем форматы: LGP, не более 10 МБ



Запустить анализ

Страница сервиса



Внутри – low-code реализация и публикация в виде веб-сервиса

Общая информация о пакете «Decomposition.lgp»

- Версия платформы: 7.2.4

Описание пакета

Пакет предназначен для комплексной обработки и анализа временных рядов, начиная с импорта данных и заканчивая их декомпозицией и визуализацией. Он включает настройку параметров, фильтрацию, восстановление пропущенных значений, группировку и заполнение данных. Пакет поддерживает анализ сезонности, трендов и стационарности, а также декомпозицию с использованием методов ARIMA, LOESS и других. Встроенные ИИ-модули (GigaChat, llama3.3) помогают в обработке и оценке данных. Пакет обеспечивает подготовку финальных наборов данных для дальнейшего анализа и визуализации, упрощая работу с временными рядами для бизнес-аналитиков.

Ссылки на внешние пакеты

- loginom_silver_kit.^3.1.4
- loginom_sklearn_kit.^3.2.1
- loginom_json_kit.^1.1.0
- __loginom_llm_kit.^1.0.0

Статистика пакета

- Количество модулей: 4
- Общее количество заметок: 50
- Общее количество узлов: 160
- Общее количество подмоделей: 41
- Общее количество узлов программирования: 3
- Общее количество узлов-ссылок: 2
- Общее количество производных узлов: 5

Список модулей

- Модуль 1. «Исходные данные»
- Модуль 2. «Первичный анализ данных»
- Модуль 3. «Декомпозиция»
- Модуль 4. «Вспомогательный модуль»

Модуль 1. «Исходные данные»

Описание модуля

Сценарий содержит 0 уровней вложенности

Сценарий описывает процесс импорта и обработки временных рядов из файлов для дальнейшего анализа. Основной упор делается на настройку импорта, фильтрацию данных и подготовку финального набора данных для использования в других частях пакета.

Основной сценарий:

- Импорт данных:** Начинается с загрузки данных из файла “Потребление электрической энергии.txt” через узел “Текстовый файл (Импорт)”. Это начальный этап, где данные поступают в систему.
- Настройка параметров:** Импортированные данные передаются в узел “Параметры полей”, где происходит настройка временных рядов. Этот этап включает обработку и структурирование данных для дальнейшего анализа.
- Фильтрация данных:** Параллельно с импортом запускается узел “Фильтр строк”, который выполняет фильтрацию данных по заданным критериям (например, по идентификатору). Этот процесс позволяет работать только с выборочными временными рядами.
- Финальный набор данных:** После обработки и фильтрации данные поступают в финальный узел, который формирует итоговый набор данных. Этот набор используется всеми остальными частями пакета для дальнейшей работы.

Логика сценария:

- Импорт временных рядов:** Основная задача — загрузить данные из внешнего источника и подготовить их для анализа. Узел “Текстовый файл (Импорт)” выполняет эту функцию.
- Настройка и фильтрация:** Узел “Параметры полей” настраивает структуру данных, а “Фильтр строк” позволяет выделить нужные временные ряды.
- Финальная обработка:** Все обработанные данные объединяются в финальный набор, который становится основой для дальнейшей работы пакета.

Сценарий не содержит вложенных подмоделей, поэтому логика распределена между отдельными узлами, каждый из которых выполняет свою четкую функцию.

Статистика модуля

- Количество заметок: 4
- Количество узлов: 3
- Количество подмоделей: 0
- Количество узлов программирования: 0
- Количество узлов-ссылок: 0
- Количество производных узлов: 0

Пример
фрагмента
отчета

ЛЕТО- ОСЕНЬ 2025

- Релиз компонентов интеграции с LightAutoML
- Релиз компонентов для обращения к LLM
- Релиз компонентов для парсинга пакета *.lgr
- Бета-версия сервиса для автодокументирования

Спасибо за внимание!