



Продуктивизация исследований: от алгоритмов и статей до внедрения и окупаемости

Савченко Максим

 @MLTools

Сбербанк, Лаборатория искусственного интеллекта

Организация коллабораций RnD подразделений



AI LAB – соединяем науку и бизнес

Оригинальные алгоритмы, модели и технологии для бизнеса



Sber AI Lab – основной хаб Сбера в области open source AI софта



Создание и сравнение рекомендательных систем

Совместно с
B2C

Платформа
AmazMe



Анализ причинно-следственных связей и A/B тестирование

Совместно с
**Блоком
Финансы**

Платформа
**SberDS
DreamML
SberCost**



Автоматическое построение моделей машинного обучения (1100+ ★ GitHub)

Совместно с
Cloud

Платформа
**SberDS и KSS
MLSpace**



Статистика энергопотребления во время выполнения кода



Оценка ESG-рисков



Представлены на Климатическом саммите ООН COP28

Sber AI Lab – основной хаб Сбера в области open source AI софта



Py-Boost

Обучение моделей методом градиентного бустинга на GPU

В2С: Рекомендации
в Сбербанк Онлайн



pytorch-lifestream

Построение векторных представлений событийных данных

КИБ, Блок Финансы, Сеть продаж, Дочерние организации



Sim4Rec

Моделирование реакции пользователя на предложения рекомендательной системы

Апробирована на данных
Дочерних организаций

Новые библиотеки
2024



TSForesight

Анализ и прогнозирование временных рядов



Stalactite

Вертикальное федеративное обучение и тестирование моделей в удобном интерфейсе



Основная Open Source библиотека в Сбере в области рекомендаций

- Предоставление SotA рекомендательных алгоритмов
- Стандартизация подходов в области рекомендаций
- Поддержка разных архитектур: PySpark, CPU (Pure Python), GPU и Multi-GPU
- Поддержка разных форматов данных: Spark, Pandas, Polars

20+

Алгоритмов

10+

Метрик



RePlay

Проприетарное
расширение

Experimental

Экспериментальная
часть библиотеки

Опубликовано 9 статей A/A* по ключевым направлениям для core-бизнеса

A* From Variability to Stability: Advancing RecSys Benchmarking Practices // KDD (ADS) <https://dl.acm.org/doi/10.1145/3637528.3671655>
Valeriy Shevchenko, Nikita Belousov, **Alexey Vasilev**, Vladimir Zholobov, Artyom Sosedka, Natalia Semenova, **Anna Volodkevich**, **Andrey Savchenko**, Alexey Zaytsev

A Does It Look Sequential? An Analysis of Datasets for Evaluation of Sequential Recommendations // RecSys (short)
<https://dl.acm.org/doi/10.1145/3640457.3688195>
Anton Klenitskiy, Anna Volodkevich, Anton Pembek and Alexey Vasilev

A RePlay: a Recommendation Framework for Experimentation and Production Use // RecSys (demo)
<https://dl.acm.org/doi/10.1145/3640457.3691701>
Alexey Vasilev, Anna Volodkevich, Denis Kulandin, Tatiana Bysheva and **Anton Klenitskiy**

A* Neural Click Models for Recommender Systems // SIGIR (short) <https://dl.acm.org/doi/10.1145/3626772.3657939>
Mikhail Shirokikh, Ilya Shenbin, Anton Alekseev, **Anna Volodkevich, Alexey Vasilev, Andrey Savchenko**, Sergey Nikolenko

A Cross-Domain Latent Factors Sharing via Implicit Matrix Factorization // RecSys <https://dl.acm.org/doi/10.1145/3640457.3688143>
Abdulaziz Samra, Evgeny Frolov, **Alexey Vasilev, Alexander Grigorevskiy and Anton Vakhrushev**

A Stalactite: Toolbox for Fast Prototyping of Vertical Federated Learning Systems // RecSys (demo)
<https://dl.acm.org/doi/10.1145/3640457.3691700>
Anastasiia Zakharova, Dmitriy Alexandrov, Mariia Khodorchenko, Nikolay Butakov, **Alexey Vasilev, Maxim Savchenko and Alexander Grigorievskiy**



LAMA: LightAutoML

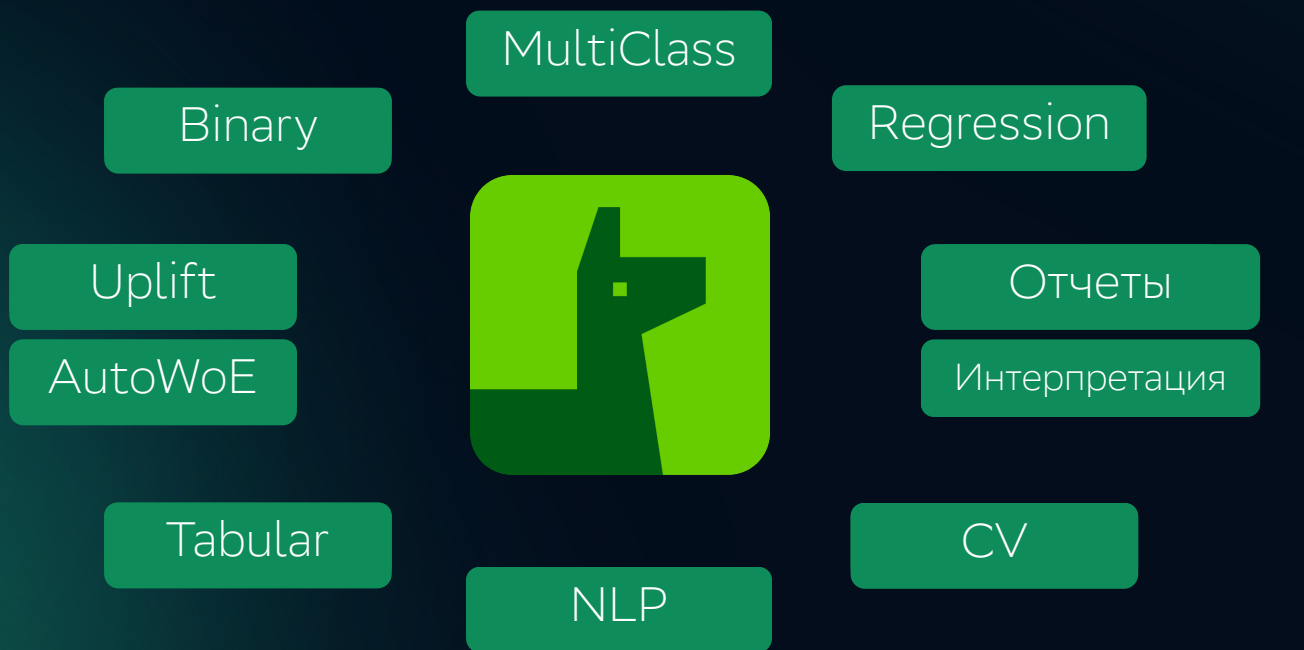
Лучшие практики участия в соревнованиях и решения задач для бизнеса

LightAutoML: AutoML Solution for a Large Financial Services Ecosystem

<https://arxiv.org/abs/2109.01528>



234K+
установок



AI решения за 1 час

Автоматически на уровне
квалифицированного специалиста

Open Source

AutoML в CLOUD, интегрирован
в ряд DS платформ Сбера

Топ 3

По признанию AutoML
команд Кремниевой долины

GPU и Spark версии

GPU-Lama и SLAMA

Модели ML:

Kaggle AutoML Grand Prix



Команда **«LightAutoML testers»** победила в международном соревновании Kaggle AutoML Grand Prix, обойдя известных конкурентов по индустрии, включая команды из Amazon и H2O с решением на основе библиотеки LightAutoML

<https://www.kaggle.com/automl-grand-prix>

Статья «LightAutoML: AutoML Solution for a Large Financial Services Ecosystem»

<https://arxiv.org/pdf/2109.01528>

Place	Team	Total Points	
1	LightAutoML testers	73	Sber LightAutoML
2	AutoML Grandmasters	72	Amazon AutoGluon
3	Robert Hatch	71	
4	Team Oxygen	57	H2O DrivelessAI (\$)
5	Dont Die Just DAI	37	H2O DrivelessAI (\$)
6	lash_fire	32	
7	SAS	28	Amazon AutoGluon Microsoft FLAML
8	Liomar	26	H2O DrivelessAI (\$)



Py-Boost – одна модель вместо 100

Подбор продуктовой полки для клиента с multilabel подходами

<https://www.biorxiv.org/content/10.1101/170803v1>

Примеры Multilabel задач

- Моделирование предложения наборов продуктов
- Повышение качества моделей в случае редких откликов в отдельных задачах

Преимущества Py-Boost

- Open Source
- Быстрые и менее затратные вычисления по сравнению с другими подходами основе деревьев (градиентный бустинг)
- Единственный алгоритм с полной поддержкой multilabel задачи без явных ограничений (по сравнению с CatBoost)



SketchBoost

Алгоритм для клиенто-ориентированного кампейнинга

NeurIPS'22 (A*)

<https://arxiv.org/pdf/2211.12858>

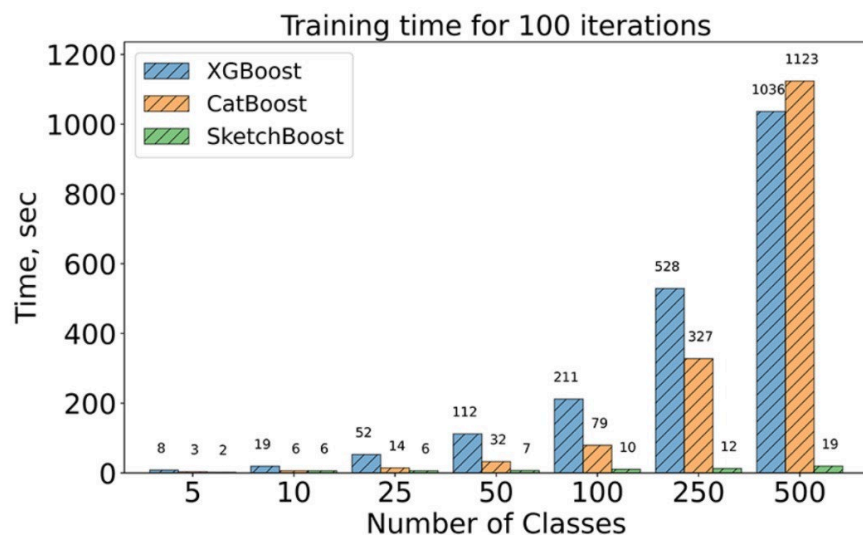


Figure 4: Training time of XGBoost, CatBoost, and SkechBoost in the same experiment as in Figure 1. Here SketchBoost uses Random Projection with sketch dimension $k = 5$. Further details are given in Appendix B.7.

Превосходство SketchBoost против классического подхода:

+4 п.п. Gini

+3 п.п. CR

Внедрение в прод СБОЛ.Про (ВСП) – замена топ-16 моделей на 1

В 16 раз

Снижение нагрузки на сервис

В 2-3 раза

Экономия трудозатрат на разработку

+1% CR

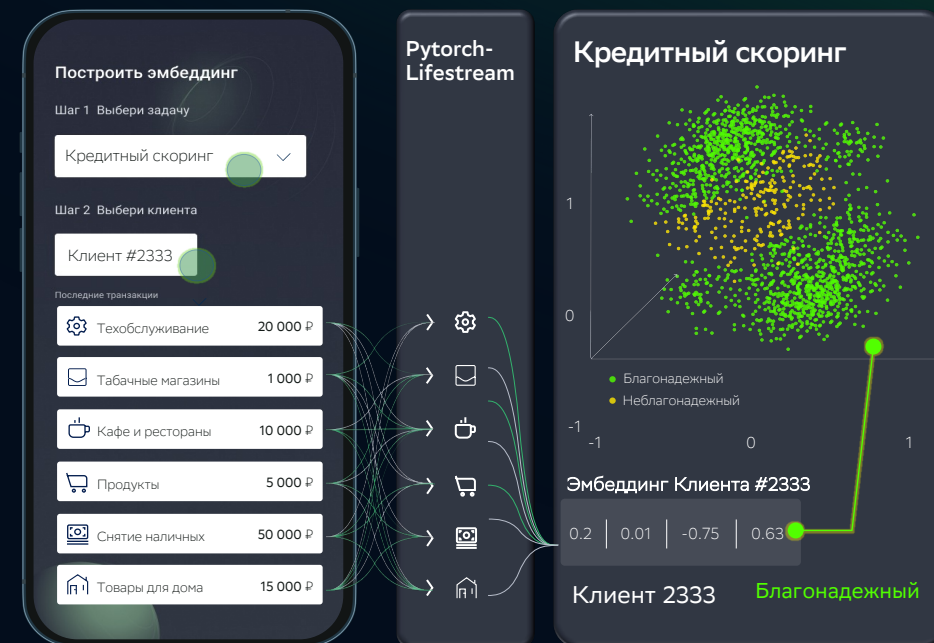
Рост конверсии



Опенсорс библиотека алгоритмов построения векторных представлений событийных данных, включающая авторские алгоритмы (например CoLES*)

Ключевые особенности

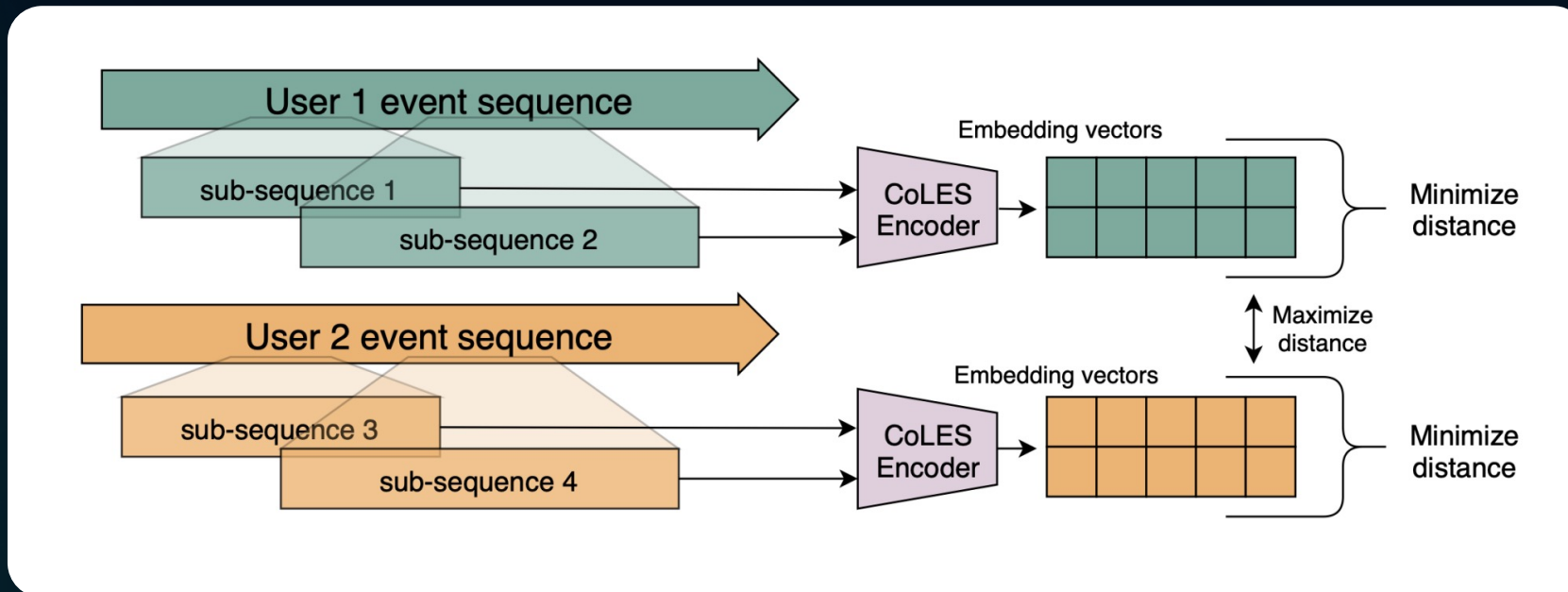
- Контрастное обучение (CoLES, CPC)
- Замена токенов (RTD)
- Предсказание следующей последовательности (NSP)
- Предсказание порядка последовательностей (SOP)
- Маскированное языковое моделирование (MLM)
- Контрастные потери, триплетные потери, VICReg и другие



CoLES

Contrastive Learning for Event Sequences with Self-Supervision [SIGMOD'22 (A*)] <https://arxiv.org/pdf/2002.08232>

- Представление всего объекта
- Явно задается понятие близости
- Сплиты подобны всей последовательности
- Нужны негативные примеры
- Нет учитывается динамика объекта



Multimodal Banking Dataset

Самый большой открытый мультимодальный банковский датасет

Датасет опубликован в открытом доступе, доступен для использования в научных работах, экспериментах итд.

Собраны и анонимизированы данные 1 млн клиентов

Задача:

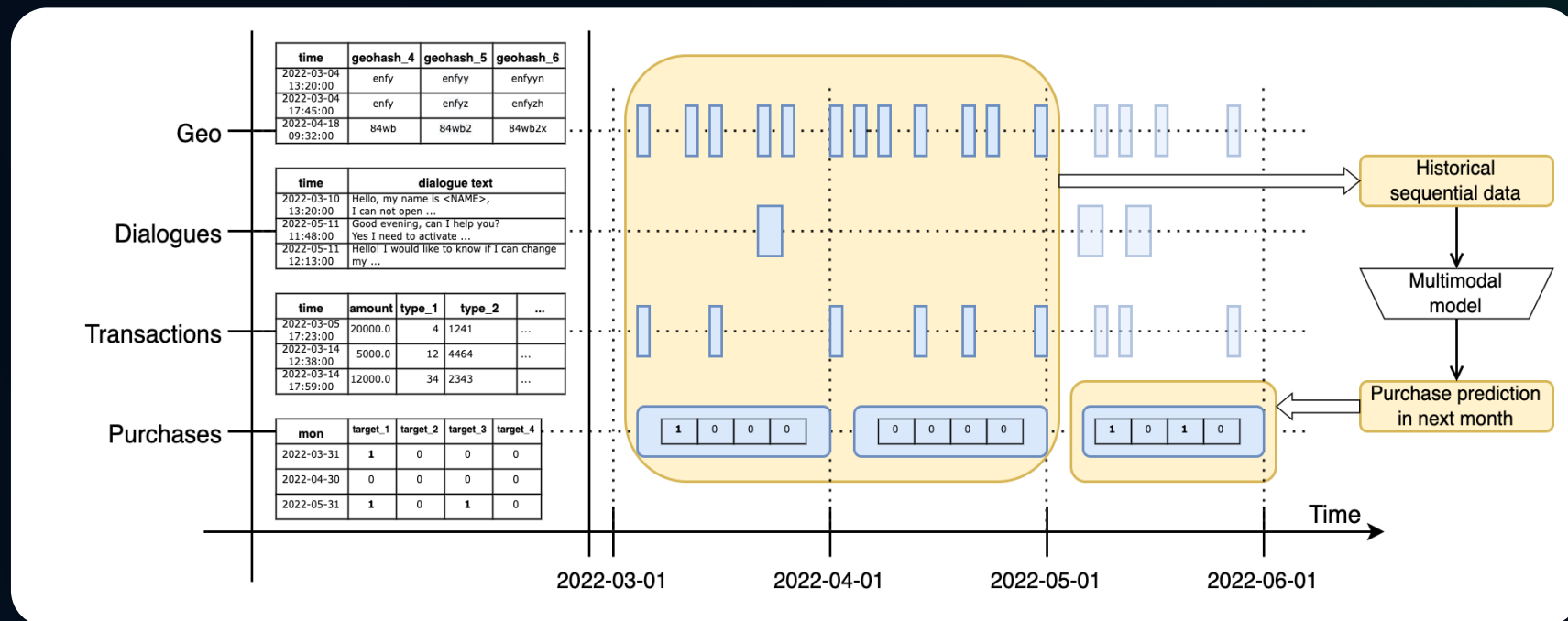
Предсказание покупки
4 продуктов за следующий
месяц

Модальности:

- Транзакции
- Диалоги
- Геострим



Mini



Инструменты анализа временных рядов

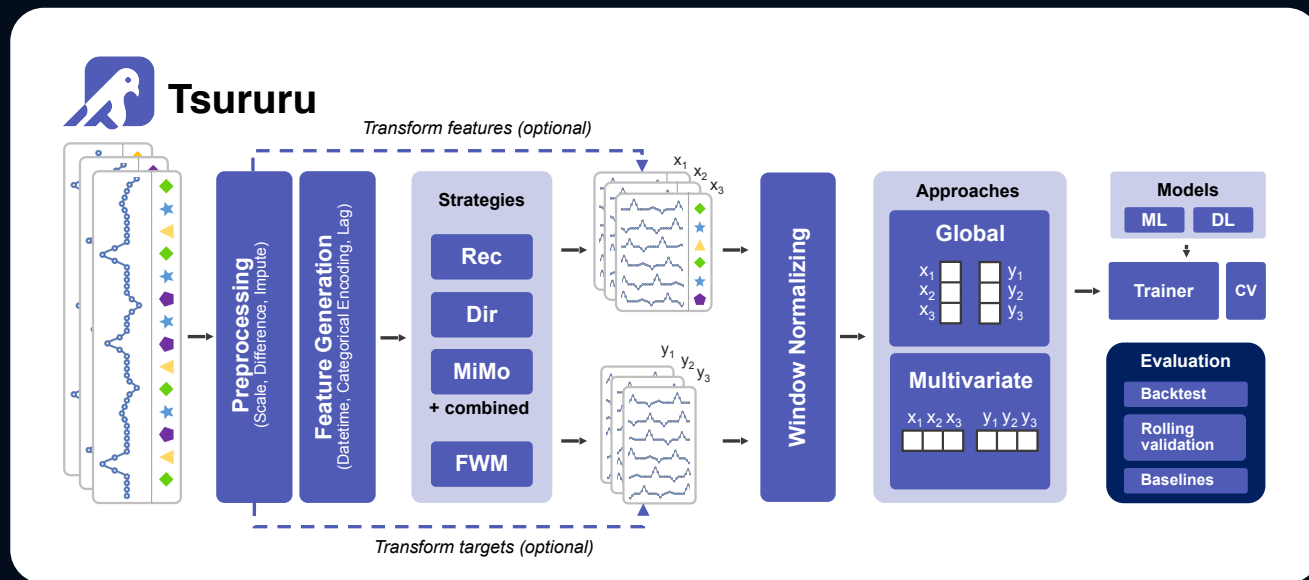
Упрощение рутинных задач и шаг в сторону Foundation Model для Рядов



TSForesight –
новая Open Source библиотека
для анализа временных рядов

- 8 Стратегий предсказания на несколько точек вперед
- 3 Подхода к моделированию большого числа рядов
- 5 Видов препроцессинга данных

Успешно протестирована
на задачах Банка



Интеллектуальный контроль возобновляемых источников энергии при кратковременной нестабильности напряжения



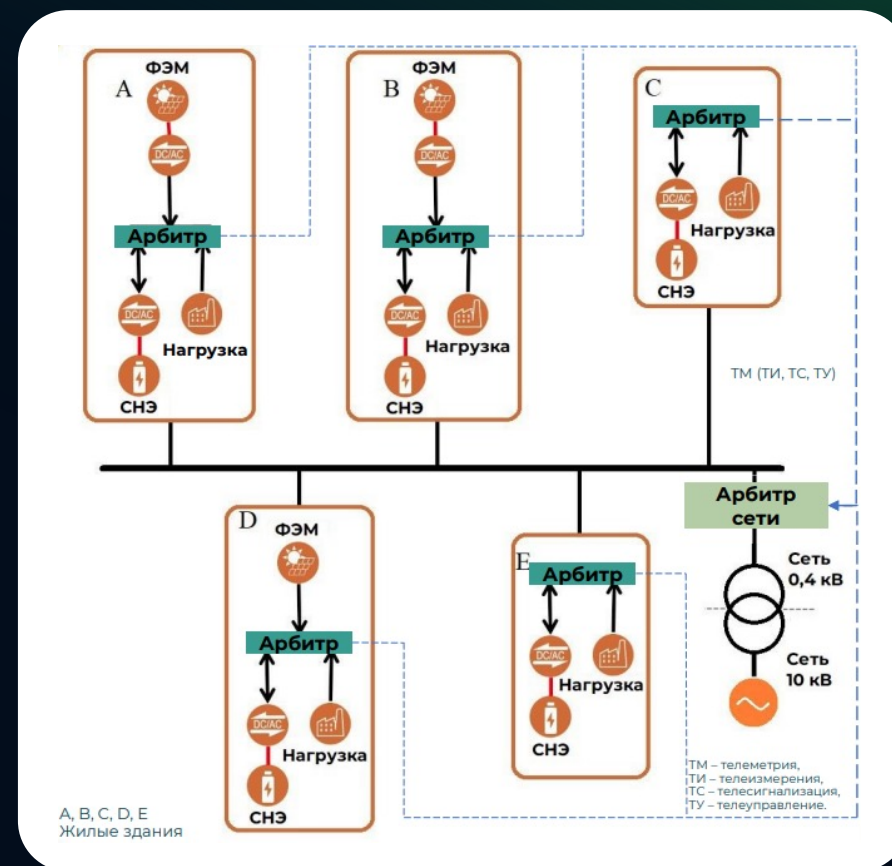
Методология

- Моделирование энергосистемы на базе PSCAD
- Анализ Временных рядов напряжения
- Регулирование работы ВИЭ на основе собранных данных

Результаты

- Повышение устойчивости энергосистемы
- Эффективное управление ВИЭ на основе анализа данных

A Novel Data-Driven Approach for Managing Renewable Energy Systems During Short-Term Voltage Instability | IEEE Journals & Magazine | IEEE Xplore



Новый метод оценки состояния изоляторов в электрических сетях с использованием моделей LightAutoML



Методология

- Выделение признаков из временного ряда утечки тока
- Реализация двухуровневого метода для оценки состояния изолятора в энергетических сетях
- Классификация и детекция воздействия загрязнения на изолятор
- Использование LAMA для реализации метода

Результаты

- ПО для мониторинга изоляторов в реальном времени
- Уменьшение затрат на техническое обслуживание изоляторов и отключение энергосистемы
- Развитие международного сотрудничества с исследователями из Сингапура и Китая

