

EESTech Challenge

REU DS CLUB & MISIS



DATA SCIENCE CLUB

Мы провели исследование и выявили следующие проблемы

1

Тракторный завод ежедневно собирает тысячи **данных различного характера**. Аналитикам приходится проделывать большую работу для их изучения

2

Необходимо решение способное своевременно уведомить о **потенциальных неисправностях** в работе машины на предприятии

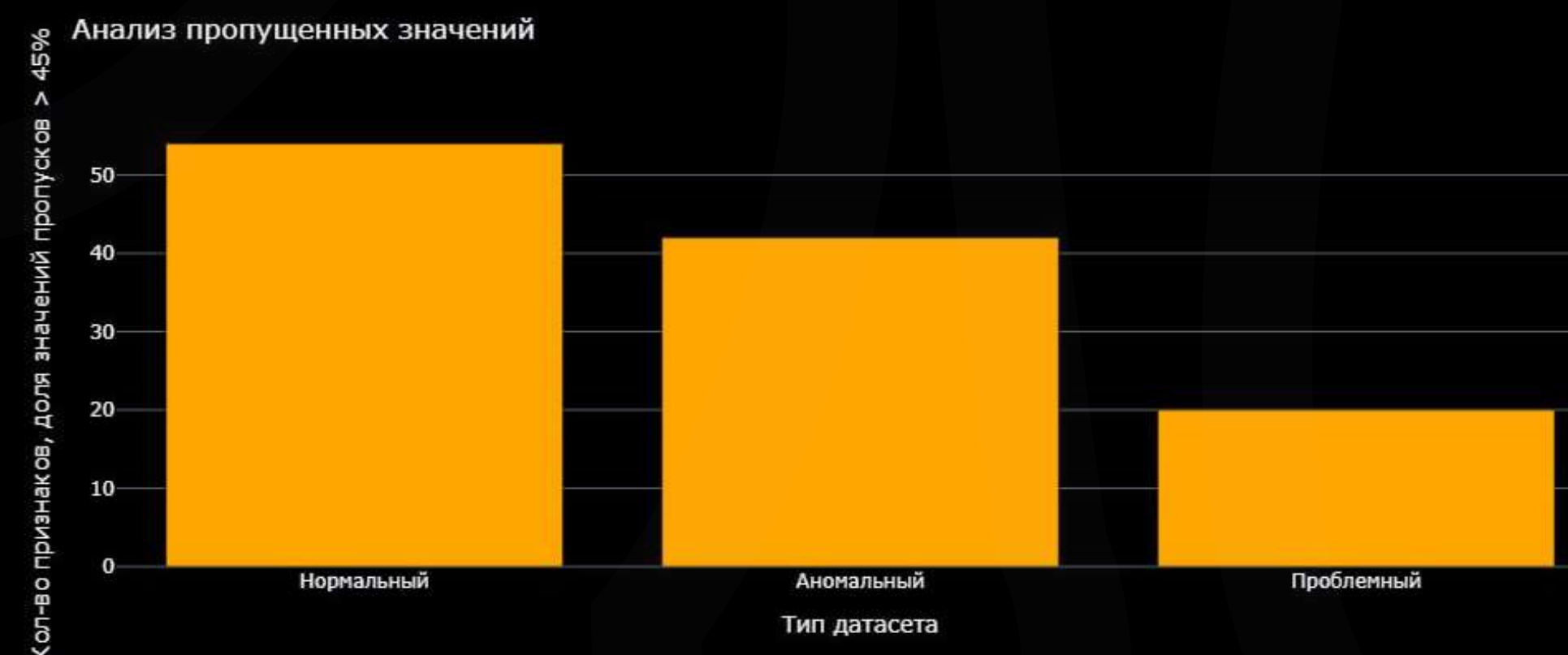
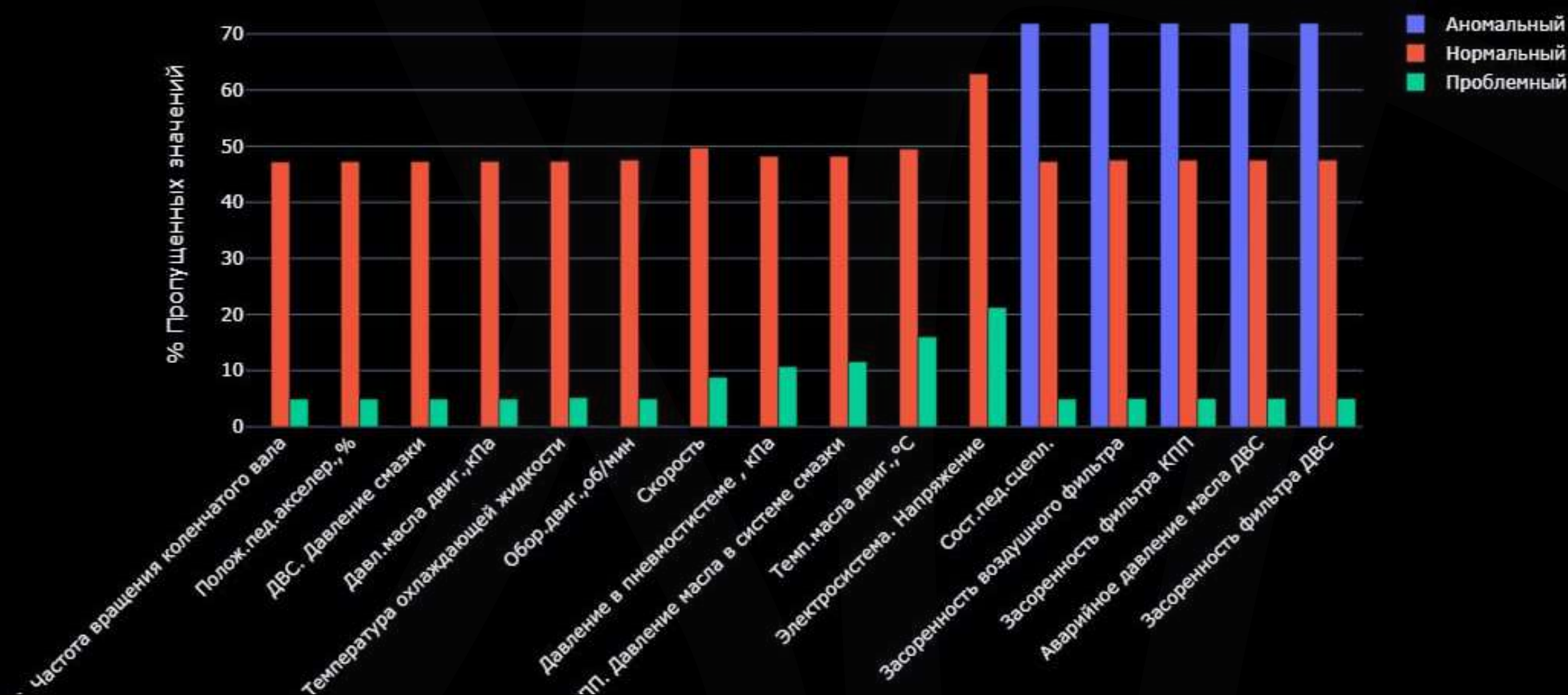
3

Решение должно иметь возможность **легко интегрироваться** в существующую систему управления предприятиями

В ходе выполнения задачи был осуществлен EDA предоставленных данных

В результате его выполнения было принято решение об удалении некоторого количества признаков

Также были произведены манипуляции над данными: созданы новые признаки, заполнены пропуски



01

Binary classification

Изначально идея состояла в том, чтобы решать задачу классификации путём прогнозирования бинарной метрики, сломается ли трактор на основе предоставленных логов или нет

02

Regression

После общения с судьями мы поняли, что важнее всего решить задачу предсказания времени поломки определенного узла

03

Forecasting

Решая задачу регрессии мы столкнулись с множеством вызовов, в итоге используя прогнозирование ключевых показателей определённых узлов техники

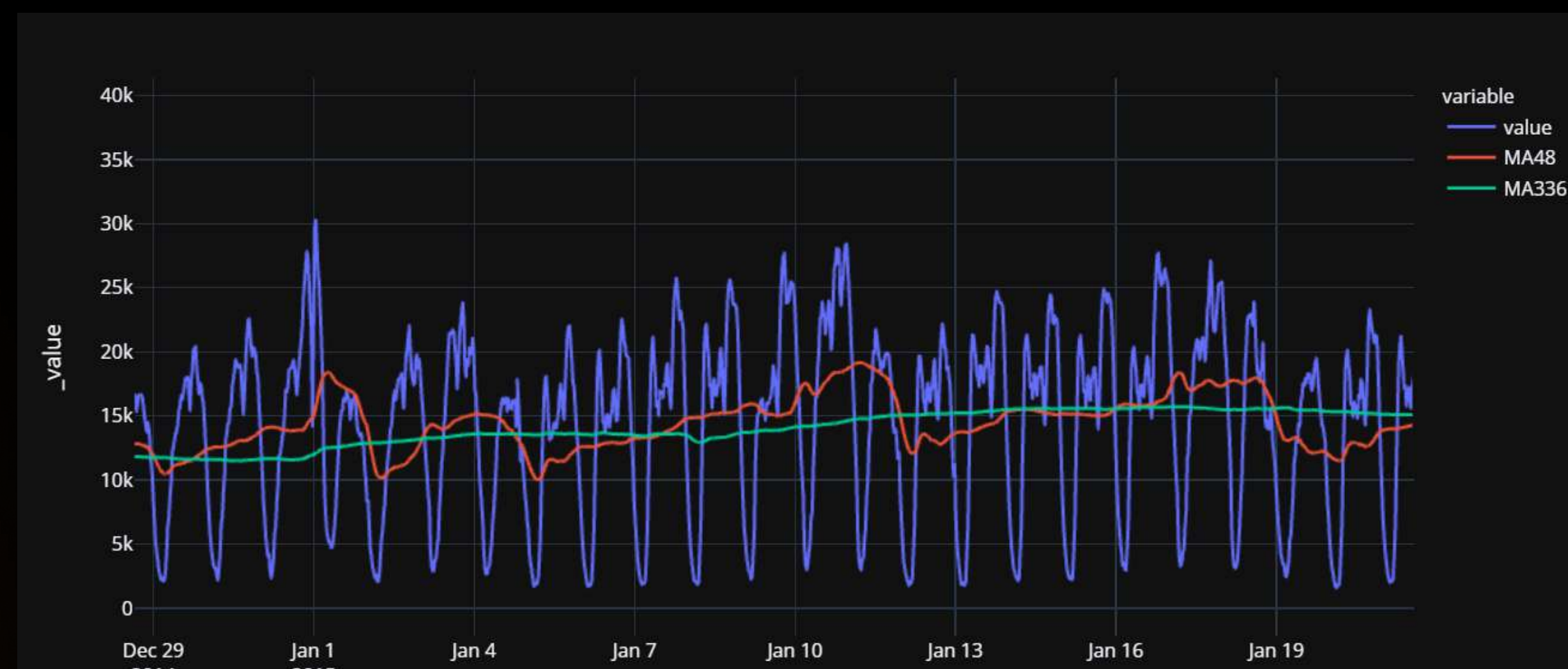
04

Prediction

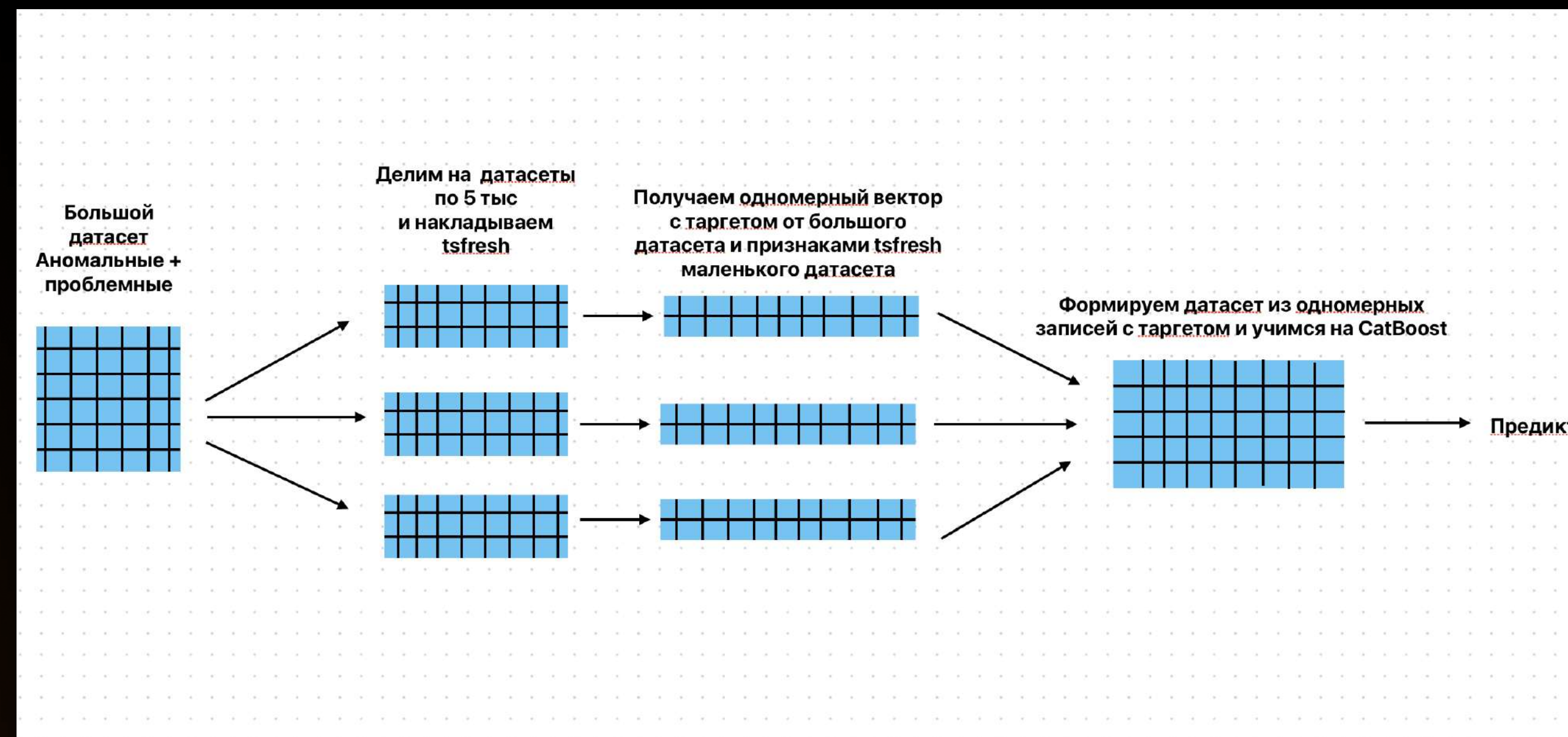
Поскольку у нас осталось немного времени в конце хакатона, мы решили попробовать реализовать изначальную идею с классификацией и обучили модель, которая на основе логов предсказывала сломается трактор или нет

Сначала данные распределялись по узлам, затем с помощью ИИ определялся ключевой признак для прогнозирования на заданный период.

Модель временных рядов Prophet была обучена на 5 признаках, каждый из которых отвечал за определенный узел, после чего по предсказаниям определялось наличие аномалий и время их возникновения.



- Уникальный подход: Объединение датасетов «Problem» и «Anomaly» в один единый набор данных.
- Разделение объединенного датасета на меньшие подмножества для детализированной обработки.
- Применение TSFresh для извлечения временных признаков из каждого подмножества.
- Объединение извлеченных временных признаков с исходными таргетами в финальный датасет.
- Обучение модели CatBoost на подготовленном датасете.





ИТММ



Решение



EESTech

log(336804182)[26-03-2024_16-06-01] 01.06-01.07.csv 8.2MB

Получить предсказания

Request successful!

	ds	engine	transmission	hydraulics	brake	electric
0	2023-06-08 10:37:24	82.204	1,153.3106	0	790.4884	27.7748
1	2023-06-08 10:37:54	82.2036	1,153.2751	0	790.4898	27.7747
2	2023-06-08 10:38:24	82.2032	1,153.2391	0	790.4907	27.7746
3	2023-06-08 10:38:54	82.2027	1,153.2026	0	790.4911	27.7745
4	2023-06-08 10:39:24	82.2023	1,153.1656	0	790.491	27.7745
5	2023-06-08 10:39:54	82.2018	1,153.1282	0	790.4904	27.7744
6	2023-06-08 10:40:24	82.2013	1,153.0903	0	790.4893	27.7743
7	2023-06-08 10:40:54	82.2008	1,153.0519	0	790.4876	27.7742
8	2023-06-08 10:41:24	82.2003	1,153.0131	0	790.4855	27.7741
9	2023-06-08 10:41:54	82.1997	1,152.9738	0	790.4829	27.774

	transmission	hydraulics	electric	engine	brake
0	Аномалий нет	Аномалий нет	Аномалий нет	Аномалий нет	10 hours



Визуализация

За основу инструмента прототипирования интерфейсов был выбран фреймворк Streamlit



Архитектура

Наше решение легко интегрируемо в любой ваш сервис за счет микросервисной архитектуры.



Backend

Были использованы следующие технологии: Python, Docker, FastAPI.



Artificial intelligence

Использование передовых средств искусственного интеллекта в решении задачи



Александр Иванов

ML

@lild1tz



Алексей Вашкевич

DS

@gasboy04



Андрей Кадомцев

MLOps

@Avenircs



Pou

Moral support

@Itstimetodrinkwater

REU DS CLUB & MISIS