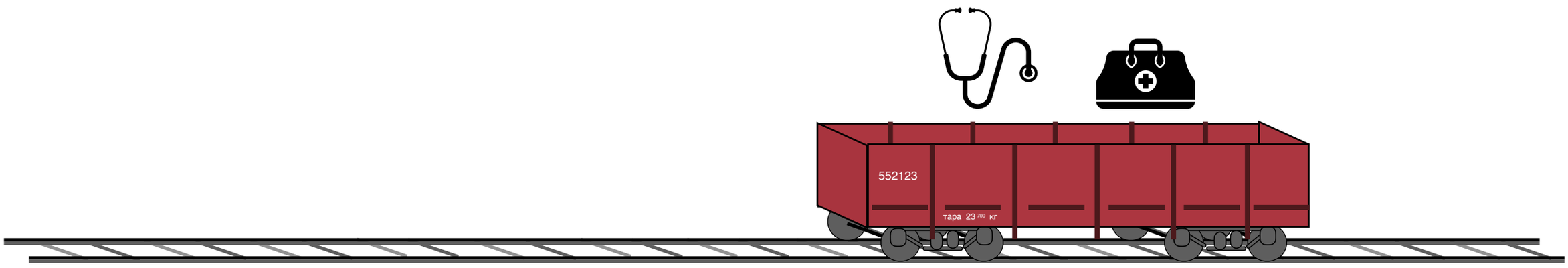
Data Wagon 2023

Трек № 2 - Чек-ап вагона

Команда fit_predict



Постановка задачи

Различные причины отправки вагона в плановый ремонт

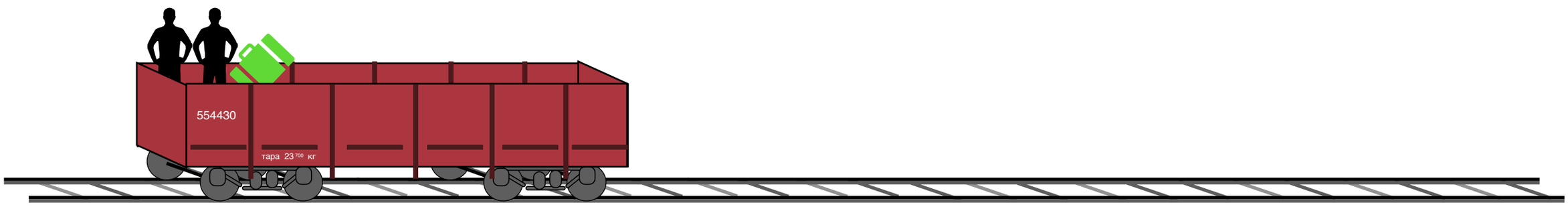
(сложные зависимости)

Текущий процесс не позволяет распределять нагрузку на ремонтное депо

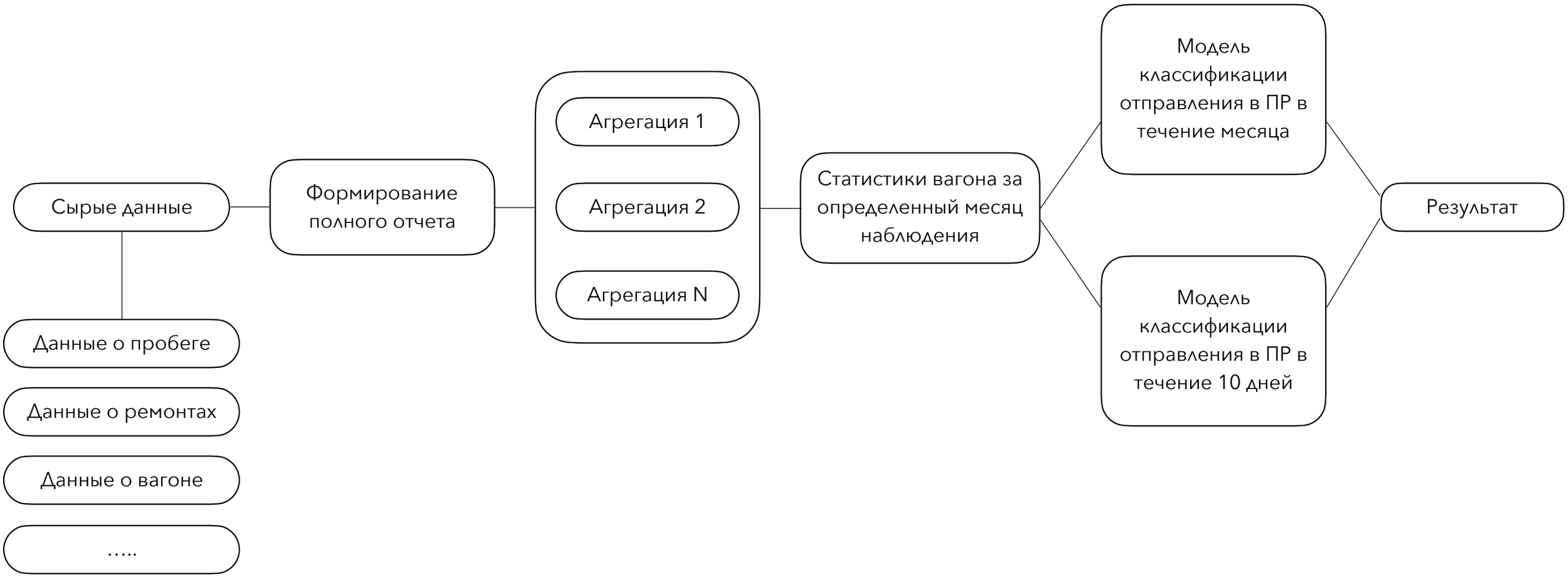
Создать модель прогнозирования даты отправления вагона в плановый ремонт (ПР)

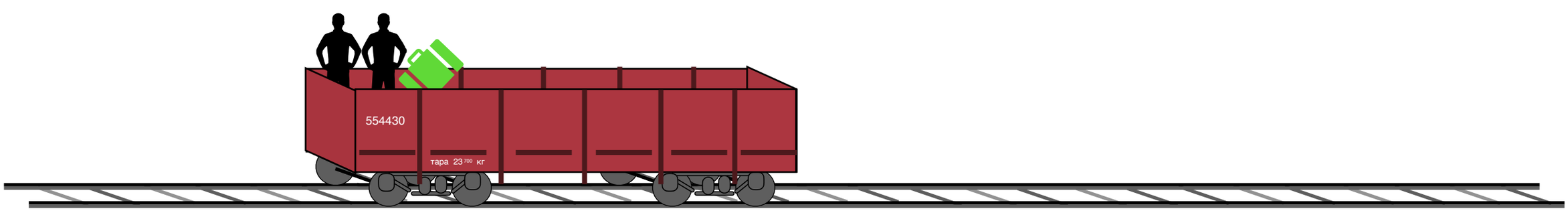
Задача прогнозирования, когда вагон отправится в ремонт, в два этапа:

1. спрогнозировать, что вагон отправится в ПР в течение месяца
2. спрогнозировать, что вагон отправится в ПР в течение 10 дней



Архитектура решения





Подготовка данных

Список вагонов, по которым известен пробег и тип владения на дату среза

Информация по дислокации

Данные по характеристикам вагона

Данные по плановым ремонтам

Данные по текущим ремонтам вагона

Данные для агрегации при наличии полного отчета ~ 7М строк

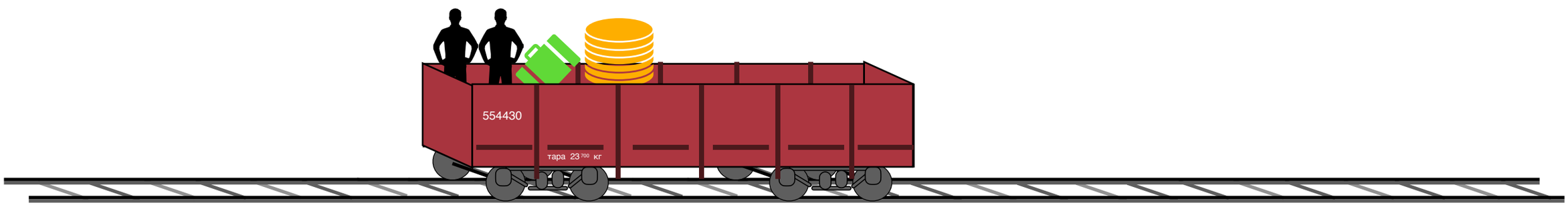
? Данные по колесным парам

? Справочник станций

Таргеты по месяцам и вагонам

Сдвиг для прогнозирования при обучении

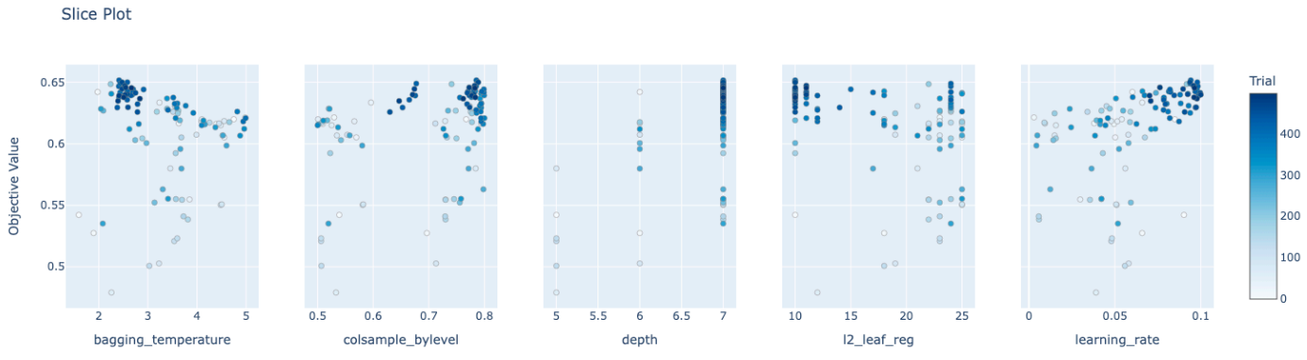
Например, имея статистики за август по каждому вагону - прогнозировать таргеты для сентября

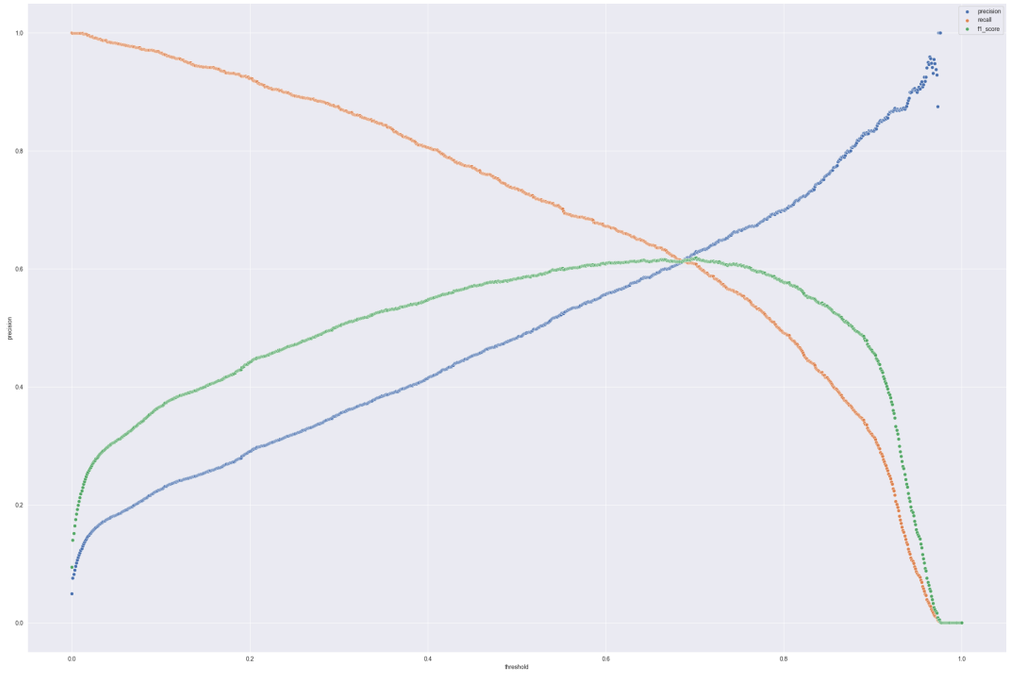


Моделирование

Y в течение месяца: **4.94 %**
Y в течение 10 дней: **1.63 %**

Метрика при обучении: **PR - AUC**
StratifidedKFold(n_splits = 5)





Поиск оптимальных параметров с помощью Optuna

Подбор порога для максимизации f1-score

Статья о PR-AUC и ROC-AUC
при дисбалансе

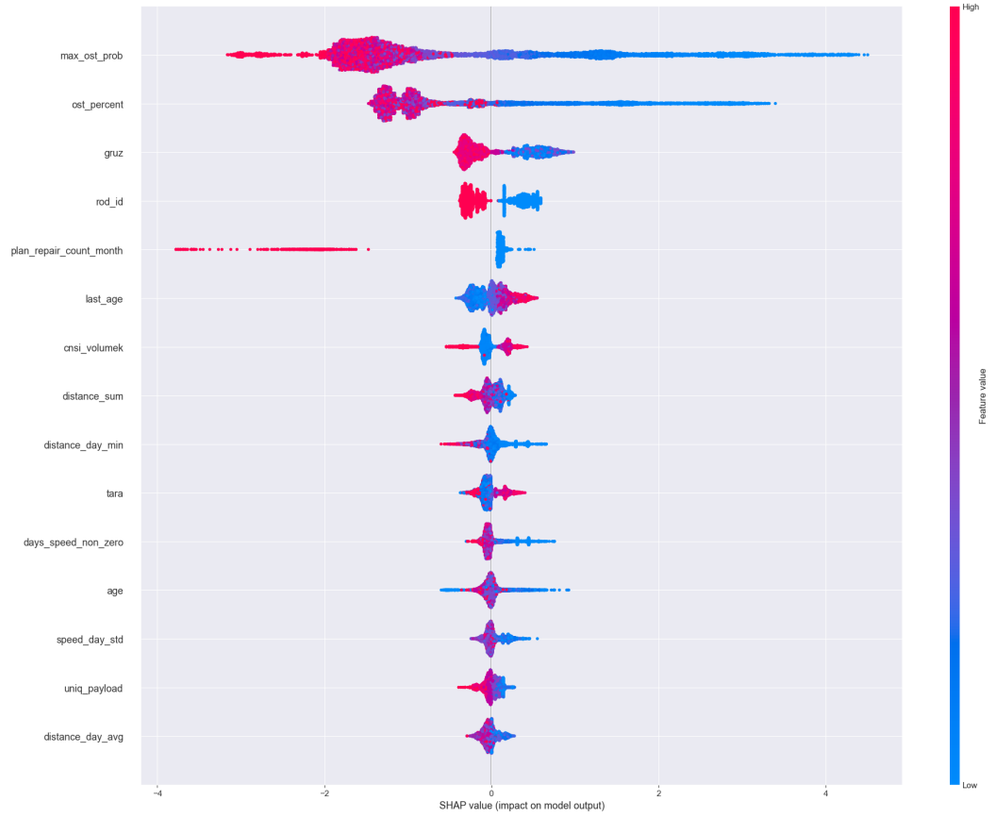
Результаты

Метрики моделей

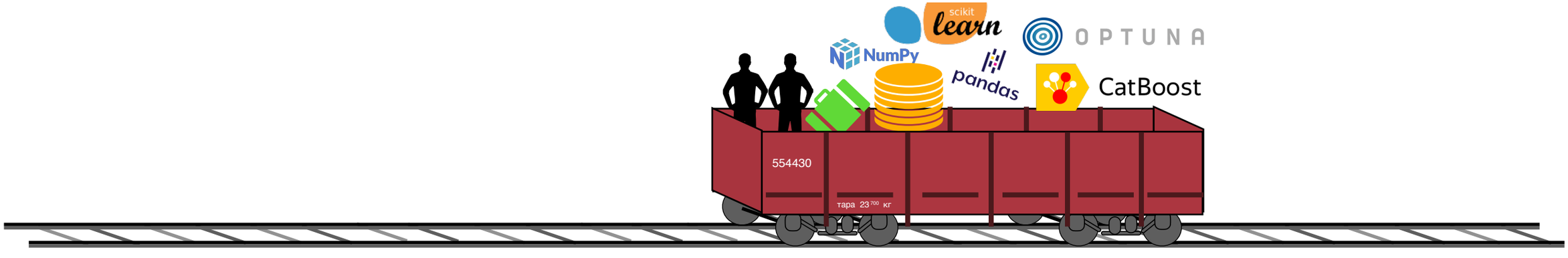
Модель	PR AUC (CV)	F1 - score (test)
Прогноз на месяц	0.655 +- 0.003	0.62
Прогноз 10 дней	0.432 +- 0.02	0.49

Этапы внедрения моделей:

1. Формирование витрин данных для моделей
2. Предсказание
    1. Batch
    2. В реальном времени
        1. Создание сервиса на FastAPI



SHAP - влияние признаков для модели прогнозирования на месяц



А что можно еще?

1. Поработать с таблицей данных о колесных парах
    1. Изменение толщины гребня
    2. Изменение толщины обода
2. Исследовать последовательности записей
    1. Последовательность ремонтов
    2. Последовательность грузов при перевозке
3. Устранить аномалии
    1. Аномалии данных пробега
4. Анализ ошибок моделей
5. Использовать > 1 модели для каждой из задач (повышение стабильности)
6. Данные о перепадах температуры / об осадках на пути движения вагонов