

/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

АО «ЕВРОСИБ СПБ-ТРАНСПОРТНЫЕ СИСТЕМЫ», 2021 – 2022

Цель работы: создание математической модели для решения задачи формирования грузовых составов на базе прогнозов вероятности попадания вагонов в текущий ремонт и программная реализация алгоритма с использованием разработанной модели.

/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

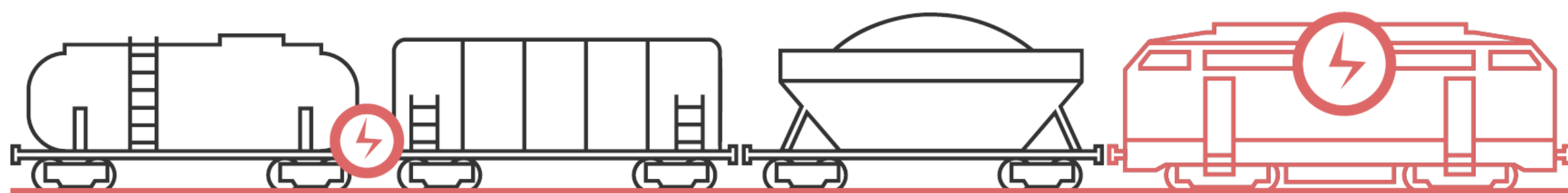
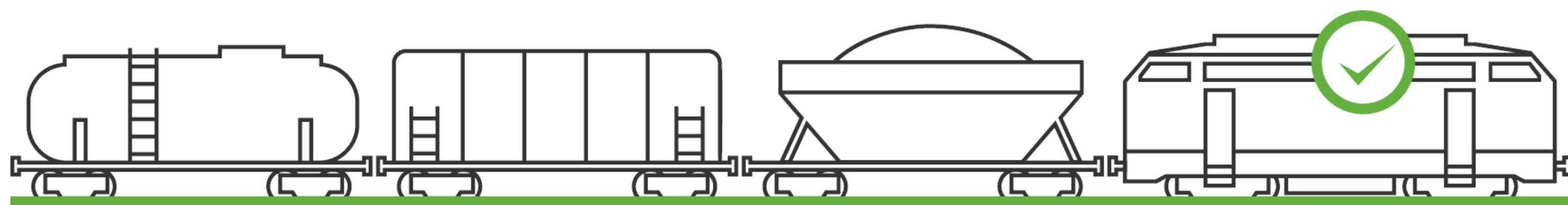
СПИСОК ПРИЗНАКОВ И ЦЕЛЕВАЯ ПЕРЕМЕННАЯ

- станции отправления и назначения;
- тип и вес груза;
- расстояние между станциями отправления и назначения;
- информация о деталях вагонов;
- история толщин гребней;
- число дней в эксплуатации вагона на дату отправления;
- данные о предыдущих ремонтах.

Требуется определить вероятность попадания вагонов состава в текущий ремонт на пути следования на погрузку или на разгрузку при выполнении грузоперевозки.

/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

ЗАДАЧА БИНАРНОЙ КЛАССИФИКАЦИИ



Задача прогнозирования сформулирована как задача бинарной классификации, и для ее решения применены методы машинного обучения с учетом различных метрик качества моделей и подбора управляющих параметров:

- логистическая регрессия;
- случайный лес;
- градиентный бустинг;
- ансамбли перечисленных моделей.

/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

ЗАДАЧА БИНАРНОЙ КЛАССИФИКАЦИИ



Особенность задачи

1. несбалансированность: количество попаданий вагонов в текущий ремонт на маршруте в итоговой выборке составляет всего около 4% от общего числа выполненных маршрутов
 2. неточности в замерах деталей;
 3. недостаточное число физических показателей.
- Это требует тщательной работы с моделями машинного обучения и выбора правильной метрики оценки качества прогнозирования.

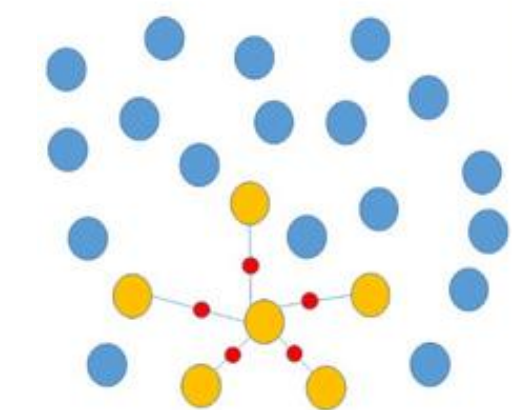
/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

ЗАДАЧА БИНАРНОЙ КЛАССИФИКАЦИИ



Методы борьбы с дисбалансом:

- использование пересэмплирования / бутстрэпа для увеличения доли объектов более редкого класса;
- субдискретизация – удаление случайных объектов многочисленного класса;
- выбор оптимальных весов целевой функции для более редкого класса;
- выбор порога округления вероятностей.



/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

МЕТОДОЛОГИЧЕСКИЙ И ТЕХНОЛОГИЧЕСКИЙ СТЕК

- алгоритм работы с большими данными;
- восстановление исходных (аномалии, пропуски);
- алгоритм генерации признакового описания объектов;
- методы машинного обучения;
- алгоритм решения задачи формирования грузовых составов на базе прогнозов попадания вагонов в текущий ремонт.



библиотеки: sklearn,
xgboost, lightgbm,
catboost, pytorch



хранение исходных
данных и результатов
расчетов

/ ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПОПАДАНИЯ ГРУЗОВЫХ ВАГОНОВ ПОЕЗДОВ В ТЕКУЩИЙ РЕМОНТ

ЧИСЛОВЫЕ РЕЗУЛЬТАТЫ: МАССОВЫЕ РАСЧЕТЫ

Порог округления = 0.5: F -мера = 0.387

	Поломки нет	Поломка есть
 Прогноз Поломки нет	9485	207
 Прогноз Поломка есть	37	77

Порог округления = 0.36: F -мера = 0.421

	Поломки нет	Поломка есть
 Прогноз Поломки нет	9453	190
 Прогноз Поломка есть	69	94

Прогнозные вероятности попадания вагона в текущий ремонт (поломки) могут быть округлены по заранее заданному порогу t . Порог округления влияет на качество модели.