

# BEST НАСЧЕТ

## ТРЕБОВАНИЯ К РЕШЕНИЮ:

Перед отправкой решения в архиве (название\_команды.zip) убедитесь, что оно соответствует следующим критериям:

1. Присутствует презентация решения в формате – название\_команды.pptx, в которой присутствует следующая информация:

- Название команды
- Имя капитана команды
- ФИО членов команды
- Подробное текстовое описание решения с обоснованием выполненных действий
- Визуализация данных и работы модели
- Результаты, полученные при вычислении указанной метрики качества.

2. Присутствует файл с прогнозом для тестовых данных в формате – название\_команды.csv

Решение отправляйте на почту [alik.nigmatullin@bmstu-best.ru](mailto:alik.nigmatullin@bmstu-best.ru) с темой [Ф] [Data Science] <Название команды>. В теле письма: обязательно должны быть указаны ФИО капитана команды.

**Все решения принимаются  
до 30 апреля, 12:59.**

# БЕСТ НАСЧЕТ

## ЗАДАЧА

В России движение грузовых поездов, в отличие от пассажирских, осуществляется по менее четкому расписанию. Перевозки планируются приблизительно и корректируются по текущей информации о статусе вагона. Продолжительность различных операций, которые совершаются над вагоном, имеет высокую степень неопределенности, которая может быть связана со сложившейся методикой работы на конкретных станциях или человеческим фактором. Мы стремимся к цифровизации бизнес-процессов и качественному прогнозированию составляющих цикла перевозки.

Основным процессом, длительность которого необходимо прогнозировать, является движение между станциями, в первую очередь потому, что именно эта составляющая жизненного цикла вагона обычно является наиболее продолжительной. Вам же мы хотим предложить задачу прогнозирования простоя железнодорожного транспорта под грузовыми операциями. Под грузовыми операциями мы подразумеваем погрузку и выгрузку различных грузов и операции, которые сопутствуют этим процессам с момента прибытия вагона на станцию и до его отправления с нее – например, расформирование поезда перед выгрузкой вагонов и его формирование после погрузки. Жизненный цикл грузового вагона может быть верхнеуровнево описан следующей схемой:

Прибытие вагона

Расформирование  
поезда

Выгрузка

Погрузка

Формирование  
поезда

Отправление

# BEST НАСЧЕТ

## ЗАДАЧА (ПРОДОЛЖЕНИЕ)

Задача состоит в том, чтобы спрогнозировать длительность процесса между событием «Прибытие вагона» и событием «Отправление».

Главным источником информации о вагоне в контексте данной задачи является железнодорожная накладная. В ней указаны сведения о грузе, грузоотправителе, грузополучателе, времени отправления, времени прибытия и другие. Соединяя две последовательных накладных для одного вагона в цепочку, можно получить временной интервал, который является целевой переменной в рассматриваемой задаче – от прибытия предыдущего рейса на станцию назначения до отправления текущего рейса.

Помимо накладных, есть дислокационные данные, которые более детально описывают жизненный цикл вагона. Некоторые события не отражены в накладных (например, выгрузка или погрузка), однако, есть в дислокационных данных – ими мы обогатили данные из накладных для того, чтобы иметь возможность разделить процесс на составляющие.

## ЗАДАНИЕ:

Необходимо разработать модель машинного обучения, способную прогнозировать длительность простоя железнодорожного транспорта на станциях грузовых операций на заданной тестовой выборке.

# BEST НАСЧЕТ

## СЛОЖНОСТИ, С КОТОРЫМИ ПРИДЕТСЯ СТОЛКНУТЬСЯ:

- В дислокационных данных могут быть ошибки и пропуски.
- Некоторые из явлений, происходящих на станции, могут не отражаться в данных, но влиять на продолжительность процессов.
- Процессы на станциях могут меняться достаточно быстро – изменение объемов погрузки/выгрузки, технологические новшества – например, появление вагоноопрокидывателей. Это, в свою очередь, приводит к заметным изменениям в продолжительности простоя.

## КРИТЕРИЙ ОЦЕНКИ:

Качество модели будет оцениваться по метрике «Доля предсказаний с ошибкой <12 ч».

## ДОПОЛНИТЕЛЬНЫЕ БАЛЛЫ:

Дополнительные баллы можно получить, если алгоритм будет иметь минимальный web-интерфейс для определения продолжительности простоя.