



Hack.Genesis
online

ПРОДУЙТЕ МЕТАЛЛ ЧЕРЕЗ DATA SCIENCE

BunchofGeeks

BunchofGeeks, 2021

Наша команда

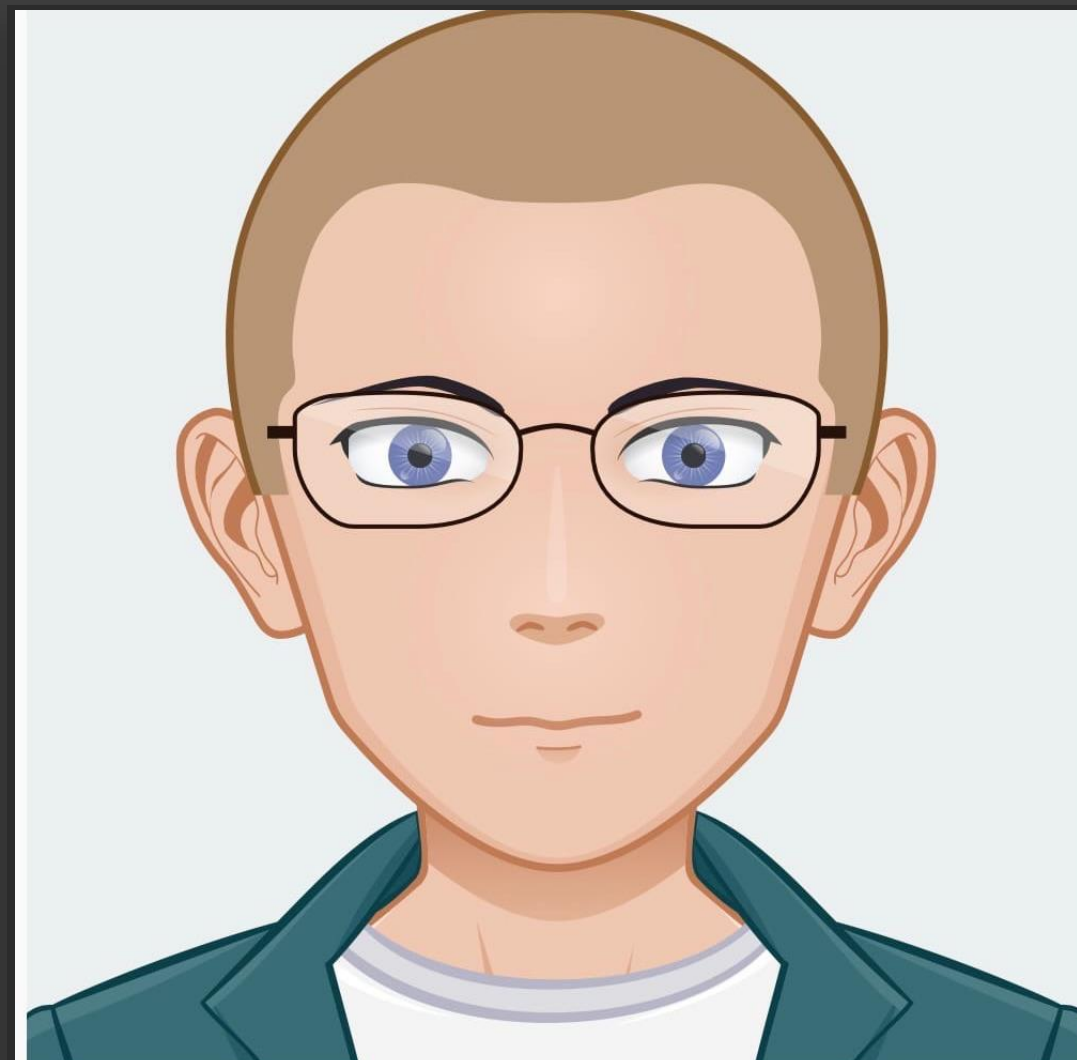


МАКСИМ ЯЗЕВ

Data Scientist

КАПИТАН

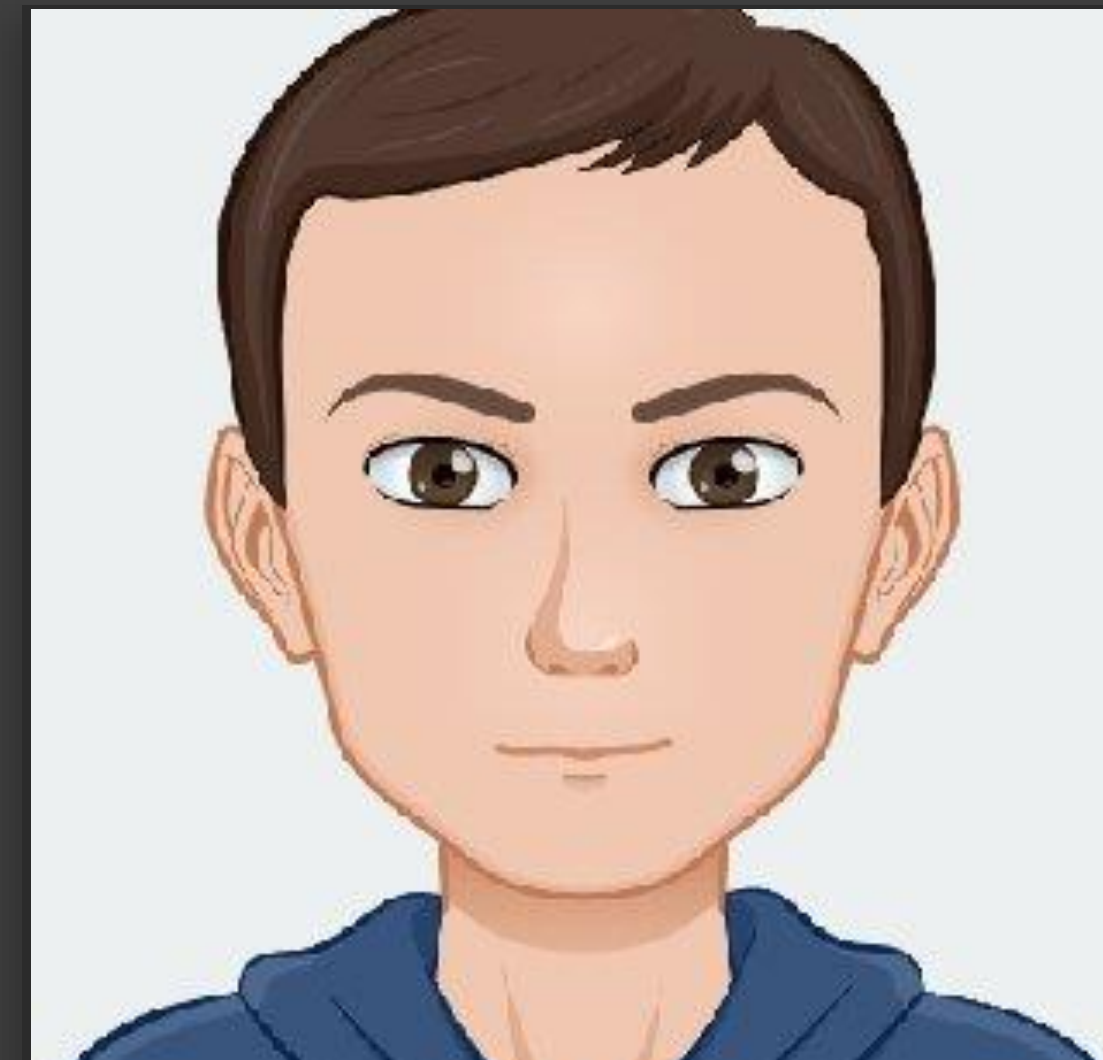
ex Sber, ex Lukoil, currently
DS @ Magnit



СЕРГЕЙ КОРОЛЕВ

Бизнес-консультант

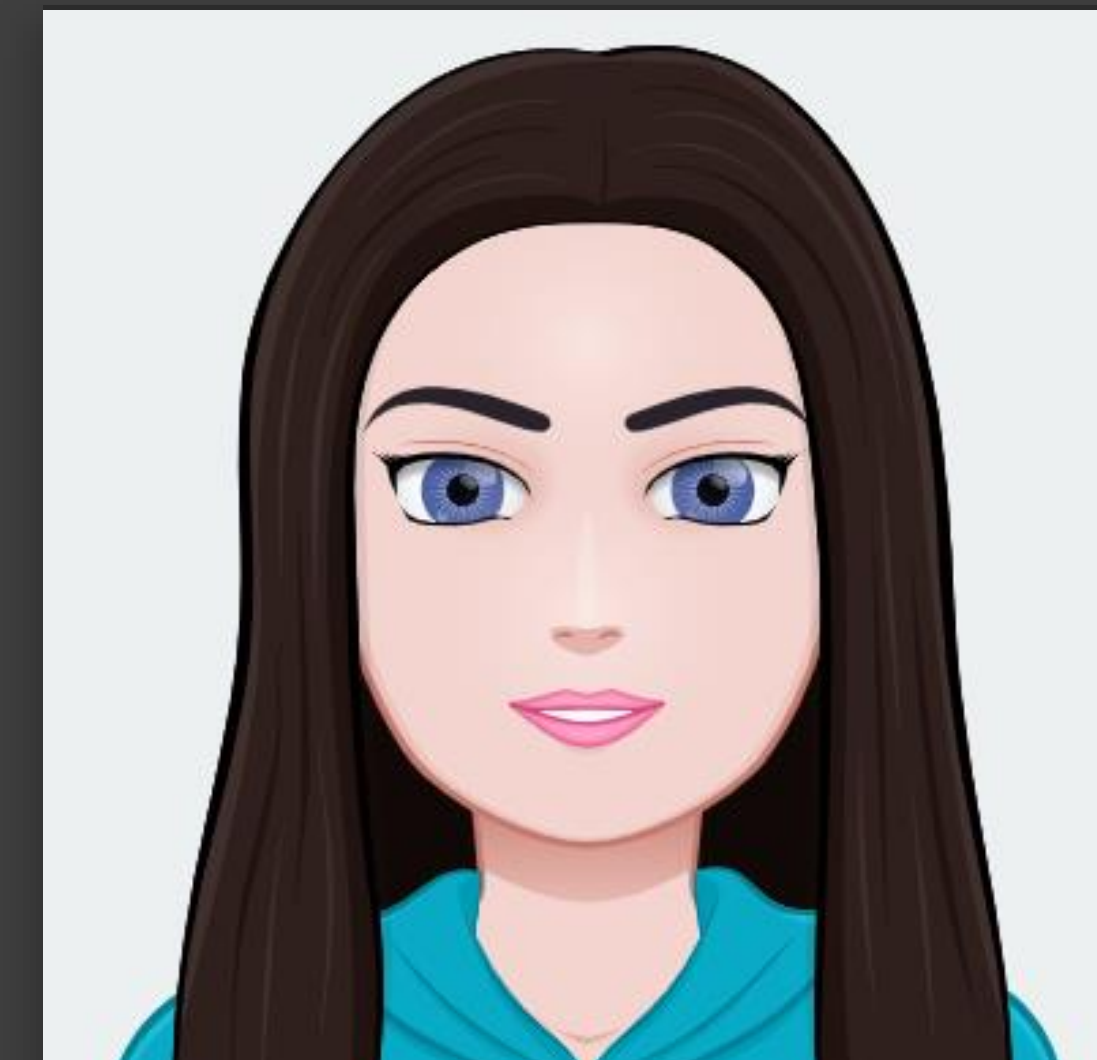
ex Sibur, currently BC @
Severstal Infocom



ДМИТРИЙ КОЗЫРЕВ

Инженер-металлург
Начинающий Data Scientist

ex Mechel, currently Freelance



РЕГИНА АХМАДУЛЛИНА

Data Scientist

*Не смогла принять
участие*

Задача

ПРЕДСКАЗАТЬ
СОДЕРЖАНИЕ
УГЛЕРОДА И
ТЕМПЕРАТУРУ
МЕТАЛЛА НА КОНЕЦ
ПЛАВКИ



Предложить
алгоритм
определения
оптимального
времени окончания
продувки

prodv



Оставляем только период относящийся к продувке по хронометражу



Суммарный и средний моментный расходы дутья на продувку



Среднее положение фурмы



Дополнительная фича - Коэффициент линейной регрессии расхода кислорода от времени с начала плавки

lom



Суммарный вес
заваливаемых ломов



Количество завалок
лома определенного
вида в течение плавки



Число разных видов
ломов в завалке



Суммарный вес лома
определенного типа в
плавке

plavki



Считаем
продолжительность
плавки в минутах



One hot encoding
категориальных фичей



Убрали дублирующуюся
плавку 511135



Логарифмируем
стойкость фурмы и
стойкость головки
фурмы

sip



Деление сыпучих на
топливо и флюсы



Число разных видов
сыпучих в завалке



Количество засыпок
определенного
добавочного в течение
плавки



Суммарная засыпка
топлива и суммарная
засыпка флюсов

chronom



Считаем
продолжительность
продувки в минутах



Выкидываем год
меньше 2021



Убираем плавки с
додувкой по t и по C

chugun



Считаем
продолжительность
между замером чугуна и
началом плавки
Отказались



Считаем VES чугуна за
базу для нормировки
других факторов:
Суммарного веса ломов,
суммарного веса
сыпучих, Суммарного
веса топлива,
Суммарного веса
флюсов,
Суммарного расхода
кислорода
Отказались



В итоге оставили эту таблицу as_is

gas (1)



Считаем баланс
 $\text{CO} + \text{CO}_2 - \text{O}_2$



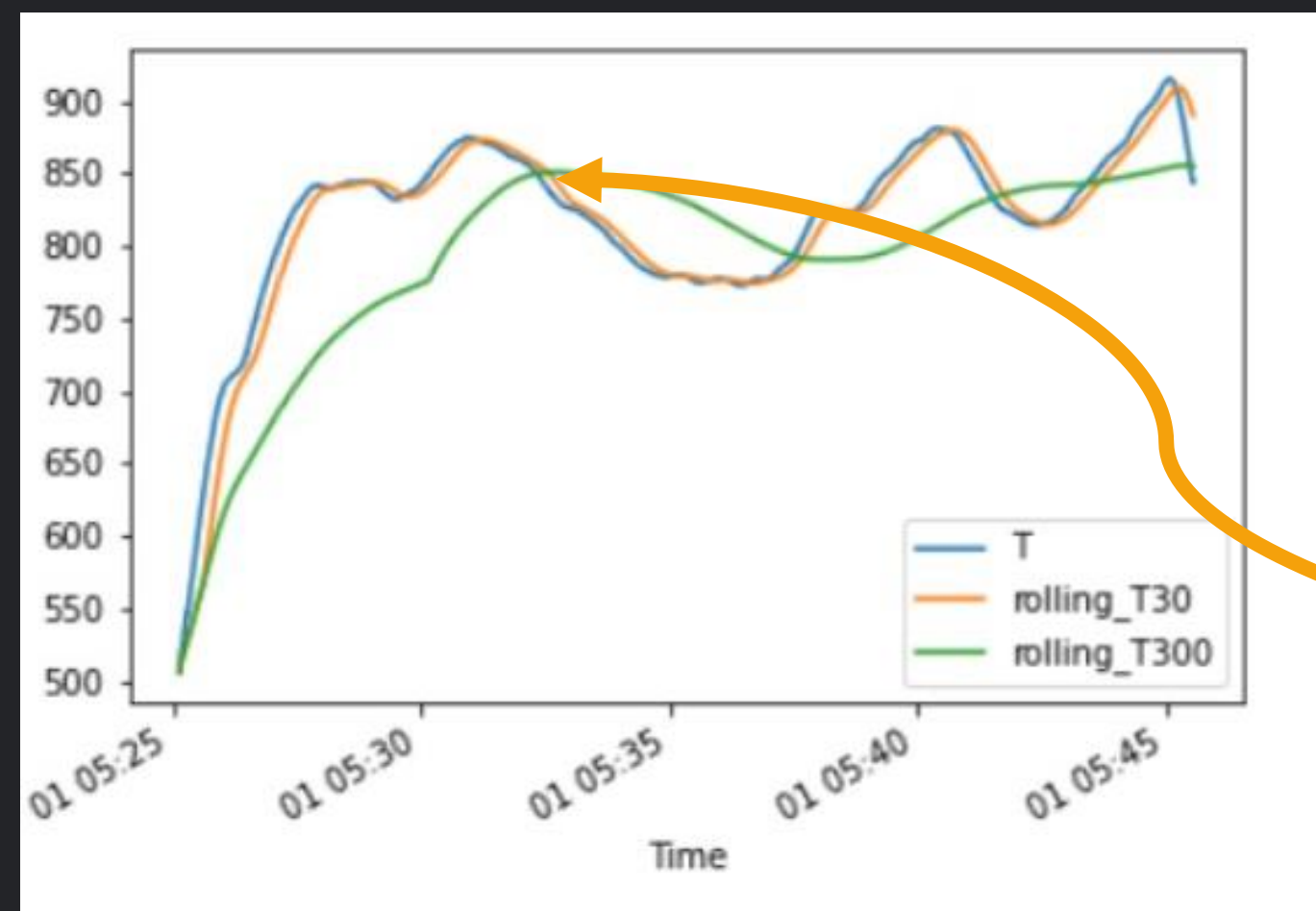
Строим сглаженные
ряды на 30 и 300 секунд
по фичам ['T', 'V', 'AR',
'balance', 'O2_pressure']



Оставляем только
период относящийся к
продувке по
хронометражу



Смотрим максимальную
из температур фурм
Отказались



**Кросс-овер и сглаженные ряды
температуры отходящих газов**

gas (2)



Считаем агрегаты min, mean, max по исходным фичам



Смотрим коэффициент регрессии давления кислорода от времени прошедшего с начала продувки



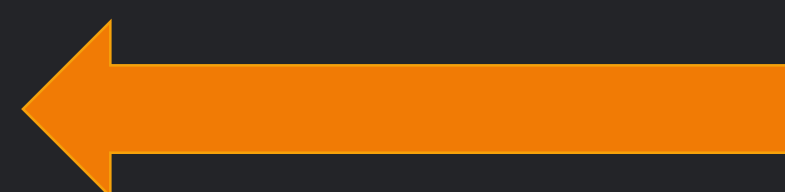
Смотрим разность скользящих средних по балансу на конец продувки



Смотрим отношение скользящих средних по аргону и температуре на конец продувки



Смотрим время от первого кроссовера температуры до конца плавки



$\text{CO} + \text{CO}_2 - \text{O}_2$: баланс

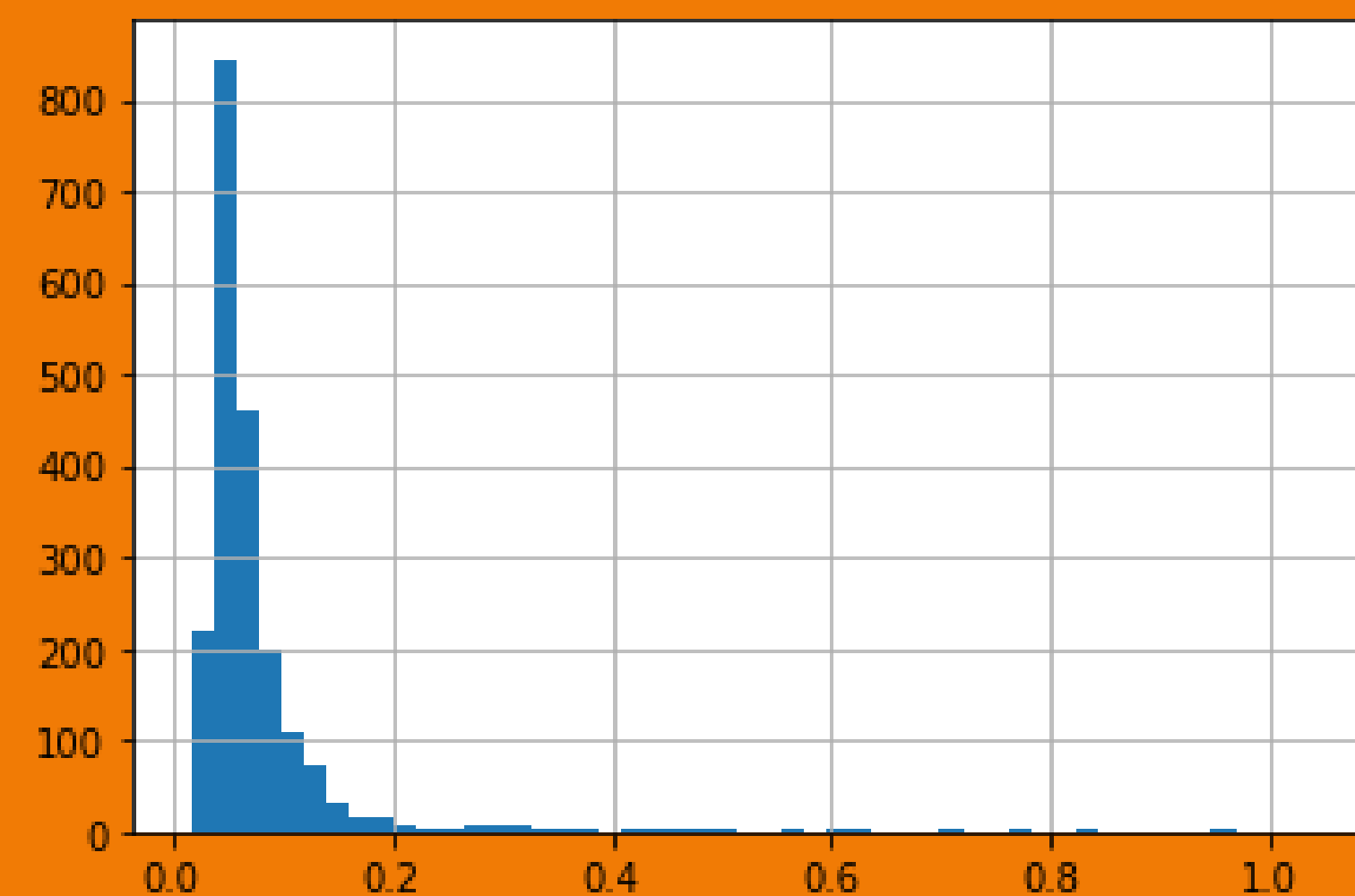
target



Удаляем пропуски для train



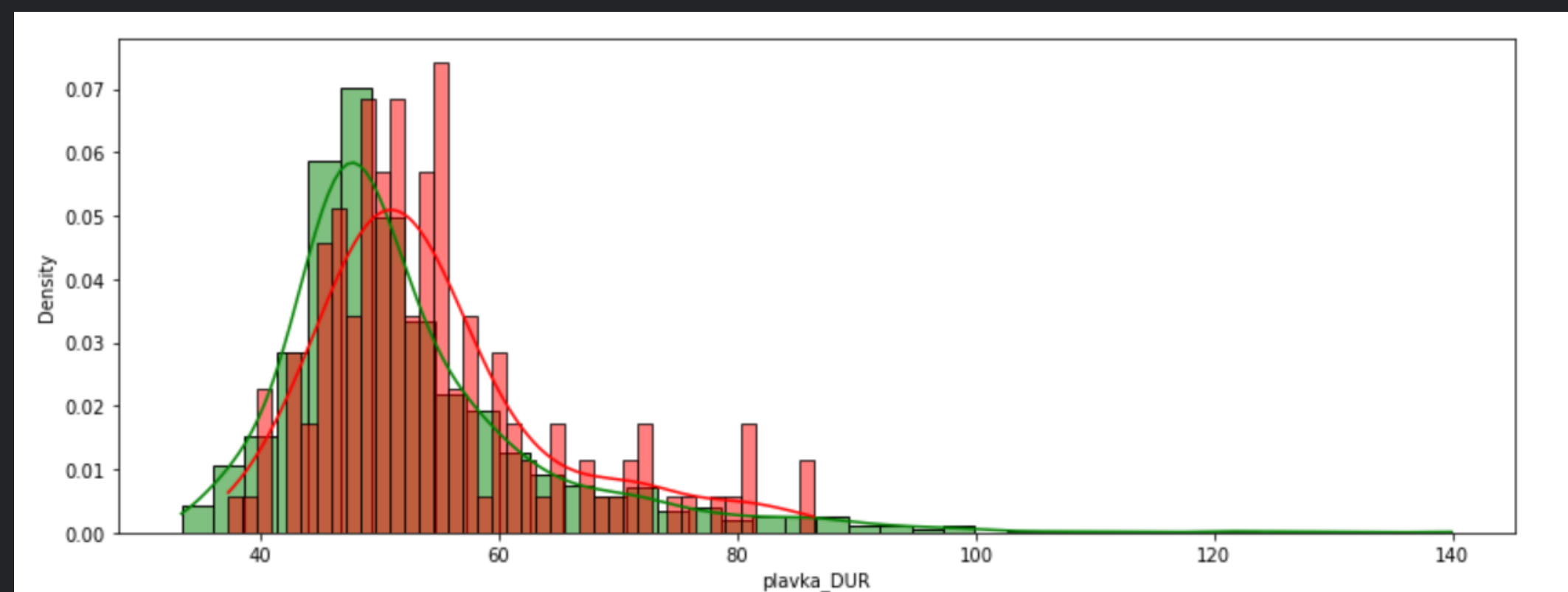
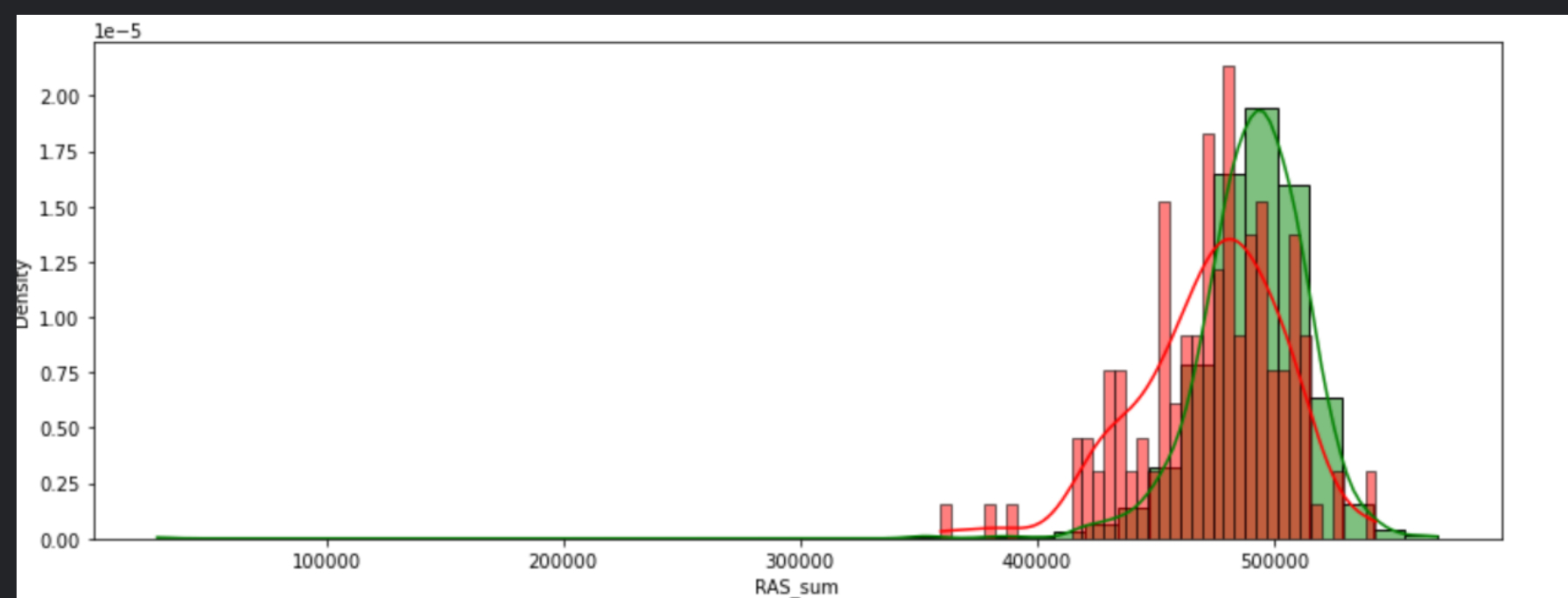
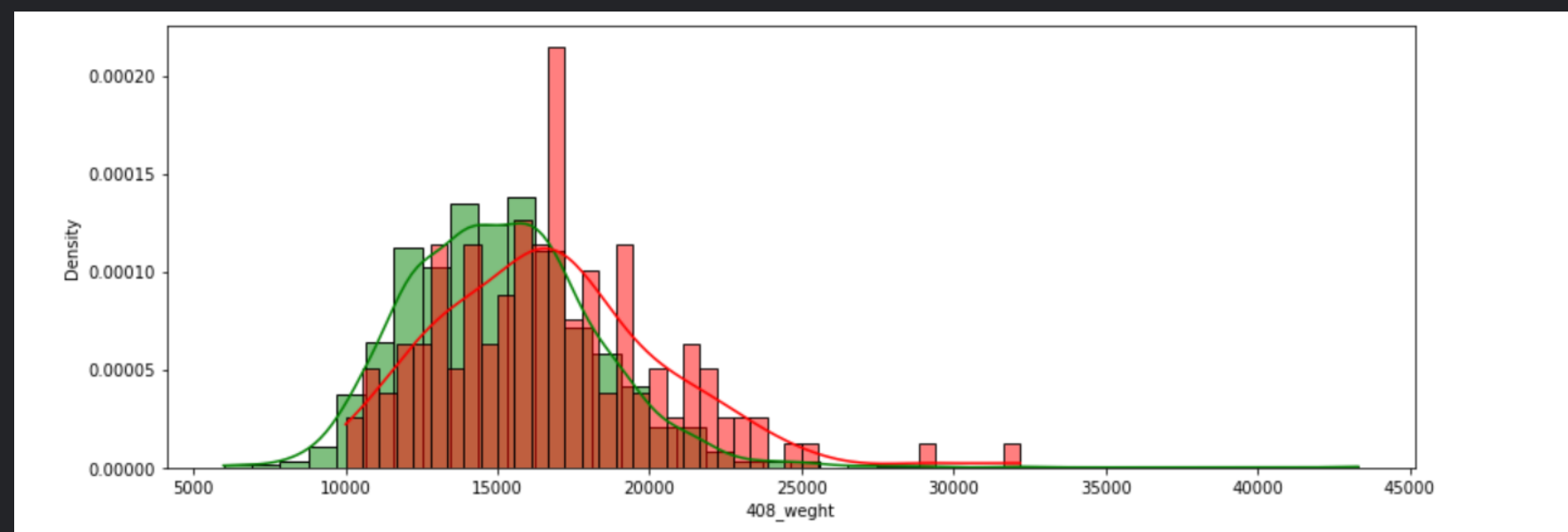
Логарифмируем углерод



Основная идея



Где мы ошибаемся



Data shift для данных по С – приводит к большим ошибкам на тесте.

*ttest_ind, alpha=0.05

Идеи по развитию модели



Использовать температуру фурмы (в train данные грязные)



Учитывать время засыпки сыпучих материалов



Модели второго уровня

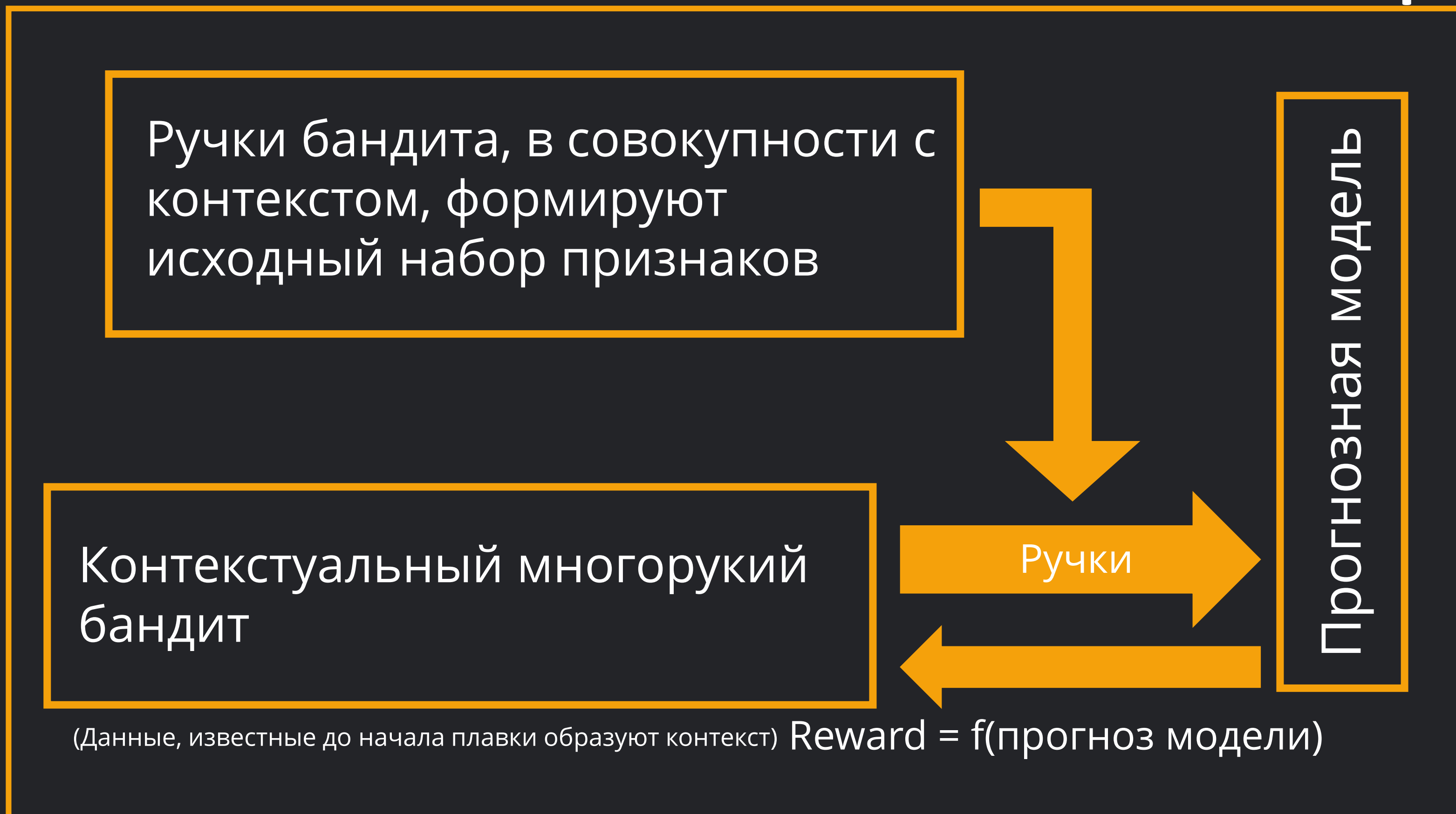


Группировать марки стали по содержанию углерода



Зависимые модели $C(TST)$, данные о предыдущих плавках и технологических операциях.

Оптимизатор



Не требователен к свойствам дифференцируемости функций



Данный алгоритм учится «онлайн», что позволяет учитывать постепенный износ оборудования

Спасибо за внимание!

