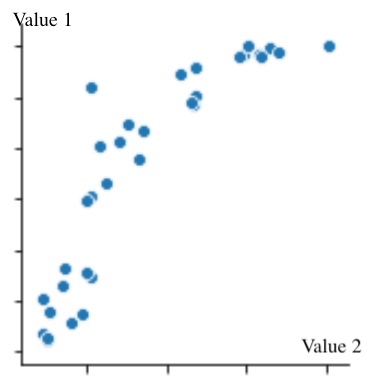
Про признаки:

У нас была первоначальная база данных, мы поверхностно просмотрели данные, а затем перешли к подготовке данных для обработки. Мы провели очистку данных, обработали текстовые и категориальные атрибуты, провели специальную трансформацию. Построили зависимости корреляции признаков между собой. Value 1 и value 2 сильно коррелировали между собой, поэтому мы создали новую метрику value = value 1/ value 2.



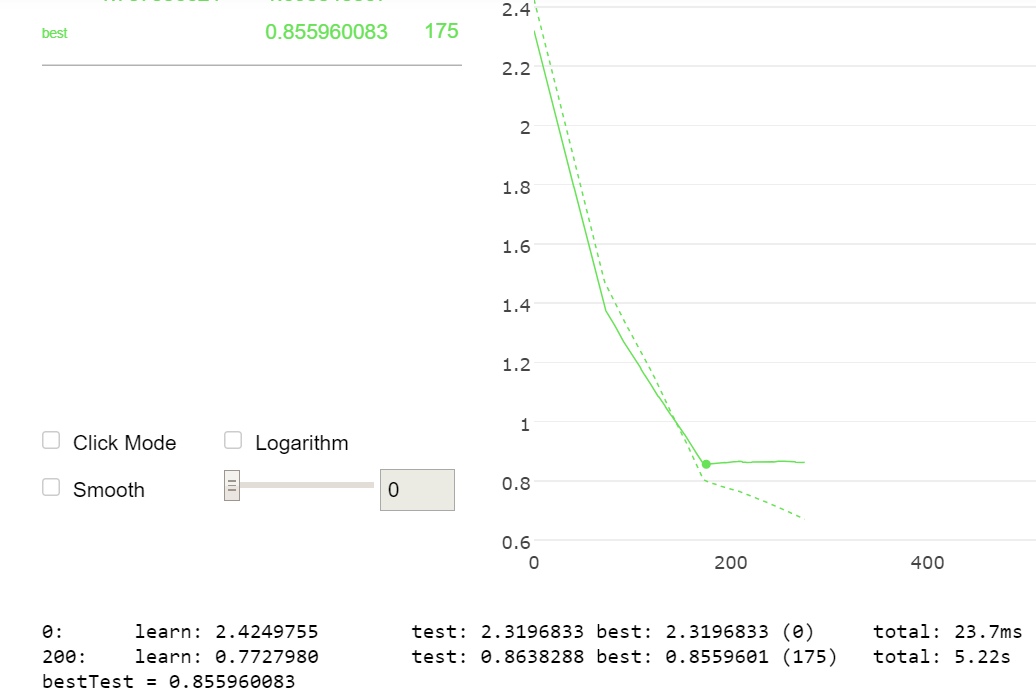
Затем мы приступили к выбору и обучению модели. Мы строили модель линейной регрессии, проводили анализ главных компонент. Так как у нас много категориальных признаков было принято решение, что будем обучать модели, используя CATBOOST.

Параллельно, мы провели сбор данных о часах работы пятерочек через яндекс api в качестве дополнительного признака. Провели парсинг и обработали данные для использования и повышения точности модели.

У нас два варианта решения, на первом мы обучались с использованием внутренних признаков, на втором без. Первый вариант: у нас есть для каждого целевого признака мы обучаем свою модель. При этом мы используем в обоих случаях модель CATBOOST.

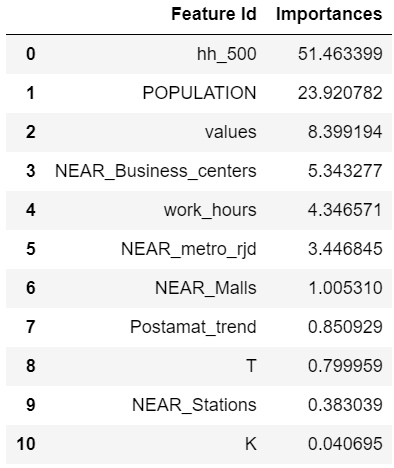
Первый вариант:

Для предсказания признака postamat daily, у нас есть 72 записи на которых можно провести обучение. 345 целевых записи для которых мы строим прогноз. Для прогнозирования целевой переменной postamat daily выбранная метрика для задачи прогнозирования – средне абсолютная ошибка(MAE).



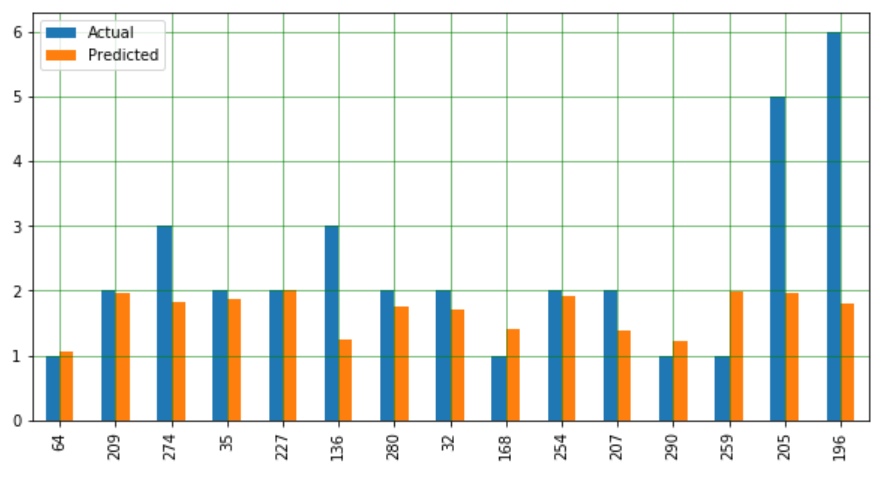
Процесс обучениея, количество итераций по икс, уменьшаем ошибку

Использованные признаки:

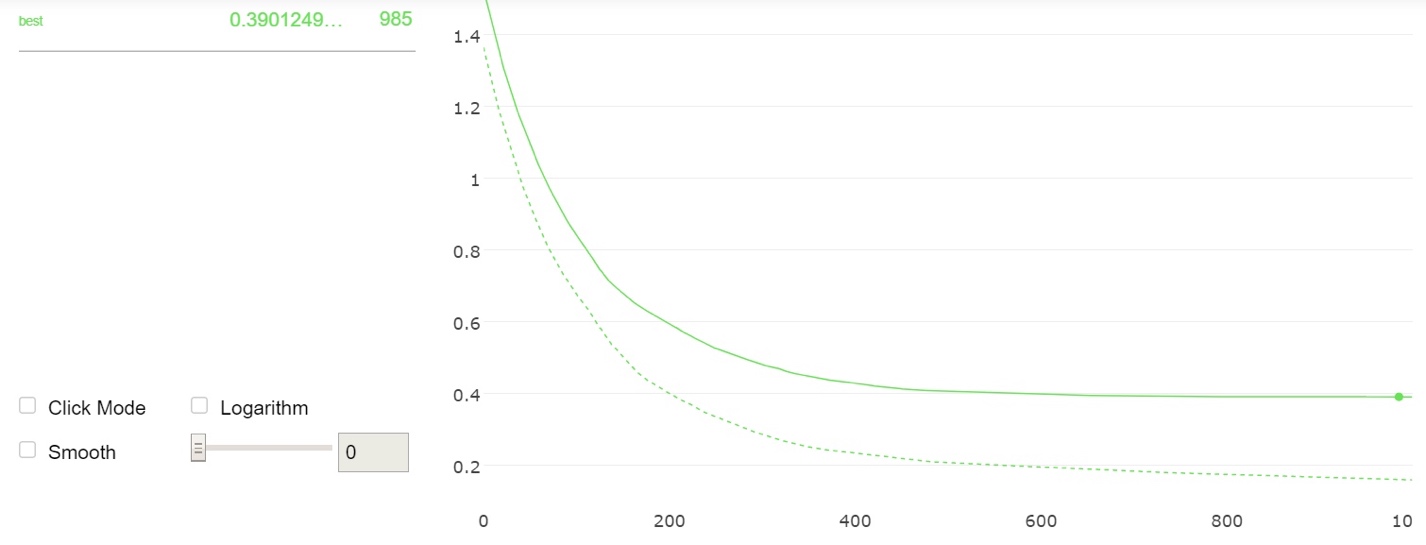


Feauture Importance

Полученные прогнозирование целевой переменной postamat daily:

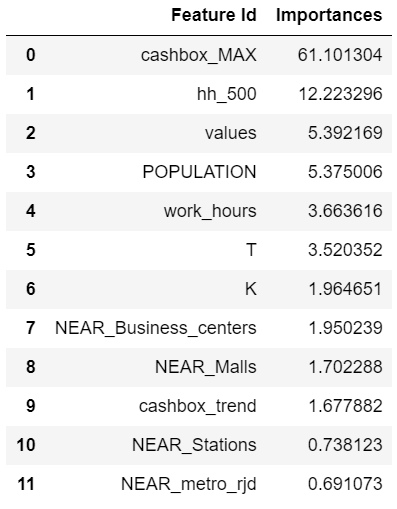


Для целевых 345 магазинов мы не знаем значений postamat trend, тк в них не установлены постаматы и нет никакой статистики. Для этого мы установили значения postamat trend самостоятельно.

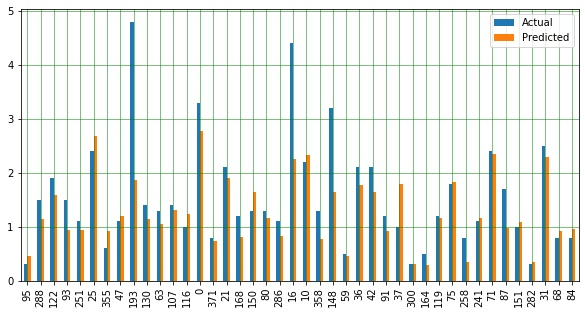
Для предсказания признака cashbox daily, у нас есть 153 записи на которых можно провести обучение. 264 целевых записи для которых мы строим прогноз. Для прогнозирования целевой переменной cashbox daily выбранная метрика для задачи прогнозирования – среднеабсолютная ошибка(MAE). 

Процесс обучение, количество итераций по икс, уменьшаем ошибку

Использованные признаки:



Полученные прогнозирование целевой переменной cashbox daily:



Мультиколлинеарность (multicollinearity) — в эконометрике (регрессионный анализ) — наличие линейной зависимости между объясняющими переменными (факторами) регрессионной модели.