

## 1. EDA и предобработка.

1. Поиск пропусков и их заполнение.
2. Удаление неинформативных признаков.
3. Очистка текстовых признаков.
4. Приведение типов.
5. Удаление дубликатов из тренировочного набора.
6. Сохранение очищенных и приведенных выборок для дальнейшей работы с моделями.



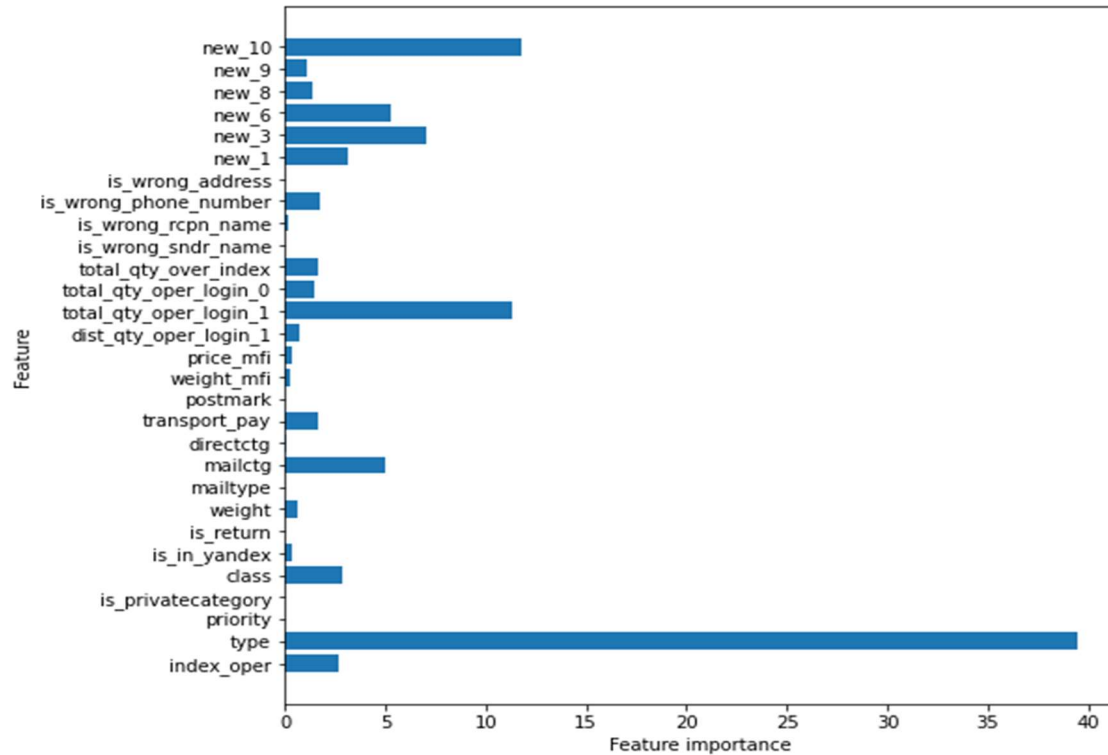
## 2. Выявление утечек.

Из описаний данных непонятно, что может являться утечкой, путем загрузки данных на лидерборд утечки были выявлены и устранены.



### 3. Создание новых признаков.

### 4. Отбор важных признаков методом RFECV (Рекурсивный отбор признаков с кросс-валидаци



5. Подготовка модели CatBoost для работы с числовыми, категориальными и текстовыми признаками. Настройка ее гиперпараметров с помощью инструмента Optuna.



6. Предсказание настроенной модели на тестовых данных и их загрузка в систему

Contact: [neolos@yandex](mailto:neolos@yandex), telegram: [@neolosik](https://t.me/neolosik)